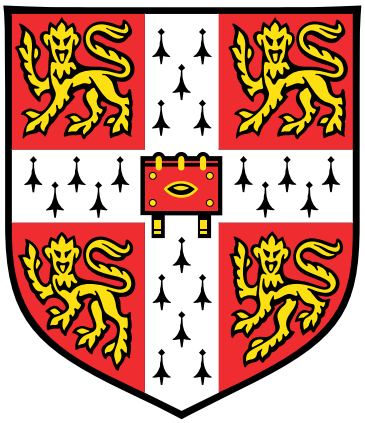
Inferring differences between networks using Bayesian exponential random graph models

Application to brain functional connectivity



Brieuc Charles Louis Lehmann

MRC Biostatistics Unit
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College

April 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Summary

Inferring differences between networks using Bayesian exponential random graph models

Application to brain functional connectivity

Brieuc Charles Louis Lehmann

The goal of many neuroimaging studies is to better understand how the functional connectivity structure of the brain changes with a given phenotype such as age. Functional connectivity can be characterised as a network, with nodes corresponding to brain regions and edges corresponding to statistical dependencies between the respective regional time series of activity. A typical neuroimaging dataset will thus consist of one or more networks for each individual in the study. Most statistical network models, however, were originally proposed to describe a single underlying relational structure such as friendships between individuals or hyperlinks between web pages. As a result, the development of these models has largely been restricted to the single network case. While one could in principle fit a single network model to each individual separately, it is not always straightforward to combine these individual results into a single group result.

In the first half of the thesis, we propose a multilevel framework for populations of networks based on exponential random graph models. By pooling information across the individual networks, this framework provides a principled approach to characterise the relational structure for an entire population. We use the framework to assess group-level variations in functional connectivity, providing a method for the inference of differences in the topological structure between groups of networks. Our motivation stems from the Cam-CAN project, a neuroimaging study on healthy ageing. Using this dataset, we illustrate how our method can be used to detect differences in functional connectivity between a group of young individuals and a group of old individuals.

In the second half of the thesis, we shift our focus to dynamic functional connectivity (dFC). Recent studies have found that using static measures may average over informative fluctuations in functional connectivity. Several methods have been developed to measure dFC in functional magnetic resonance imaging (fMRI) data. However, spurious group differences in measured dFC may be caused by other sources of heterogeneity between people. We use a generic simulation framework for fMRI data to investigate the effect of such heterogeneity on estimates of dFC and find that, despite no differences in true dFC, individual differences in measured dFC can result from other (non-dynamic) features of the data. We then add a natural and novel extension to our multilevel framework by inserting time windows as an intermediate level between time points and subjects. Using magnetoencephalography data from the Cam-CAN study, we apply our method to detect differences in time-varying connectivity between a young group and an old group.

To Paula

Acknowledgements

This thesis would not have been possible without the financial, academic and personal support from a wide range of sources. I am very grateful for funding from the Medical Research Council and the Cambridge Philosophical Society. I am also thankful for travel funds from COSTNET, ISBA, and Trinity College, Cambridge, which permitted me to attend a number of highly rewarding conferences and workshops.

The patience and guidance of my supervisors Simon White, Rik Henson and Linda Geerligs proved invaluable throughout the course of my PhD. The MRC Biostatistics Unit provided a rich environment to explore new ideas; I am especially indebted to Rob Goudie, Fiona Matthews and Stephen Hill for their useful suggestions. I would not be where I am now without the inspiration of my teachers, particularly Vadim Myslov for first sparking my interest in statistics, John Aston for introducing me to statistical neuroimaging, and Ton Coolen for his advice and encouragement.

While progress towards the completion of this thesis was patchy at times, the support from my family and friends was unfailing. I am grateful to Roxana, Mycroft, Micah, Nelson, Esma, Fabian, Ruth, Sofia, Max, Drakšo, Wiktor, Maciek, Filip, and Ravi (amongst many others) for all the great times we spent together, and I look forward to sharing many more. Finally, to Paula, thank you for your companionship and unwavering support.

Contents

Chapter 1

Introduction

The human brain is a vastly complex system that, at multiple spatial scales, can be usefully characterised as distinct units in interaction with each other [22]. At the smallest scale, the units are neurons while at the other end of the scale, the units are entire brain regions. Recently, there has been a shift in cognitive neuroscience away from a modular approach (focus on units) to a network approach (focus on interactions) [20]. Mounting evidence indicates that treating brain regions as independent processors is misguided [49]; gaining insight into cognitive function requires an understanding of how distinct brain areas work in conjunction [50].

Different neuroimaging modalities allow researchers to probe different types of brain connectivity: structural, functional, or effective. Structural, or anatomical, connectivity refers to the physical links between distinct brain regions and can be measured, for example, using diffusion tensor imaging [4]. In contrast, functional and effective connectivity are both based on neural activity. This neural activity can be measured, for example, using functional magnetic resonance imaging (fMRI) [120] or magnetoencephalography (MEG) [21]. Functional connectivity, the focus of this thesis, corresponds to the statistical interdependence of neural activity in different regions. Effective connectivity refers to the influence of brain regions on one another. While functional connectivity is a notion of undirected dependencies, effective connectivity aims to capture the causal effects of neural activity between brain regions.

Network theory provides a natural framework to model brain connectivity. The brain units correspond to the nodes of a network, or graph, and the interactions to the edges between the nodes. Recent developments in network theory [81, 103] have provided a wide range of tools for neuroscientists to study the brain. These have largely consisted

of various network measures that quantify different aspects of the brain's connectivity structure [121].

The goal of many neuroimaging studies is to better understand how brain connectivity changes with a given phenotype such as age [52], cognitive function [155] or disease status [5, 32]. In the context of network theory, this amounts to a problem of network comparisons. To assess group-level variations in brain connectivity, it is therefore crucial to develop statistical methods for the inference of differences in the topological structure between groups of networks. Given the complex structure of network data, this is a challenging task and the development of such methods is still in its infancy.

In the remainder of this Introduction, we review functional connectivity (Section 1.1), the use of networks in neuroimaging (Section 1.2), and introduce the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study (Section 1.3).

1.1 Functional connectivity

Neural activity can today be measured using a number of neuroimaging technologies, including functional magnetic resonance imaging (fMRI), and magnetoencephalography (MEG). In the modular approach to neuroscience, a typical neuroimaging study would aim to detect which brain regions are more active when performing a particular task [63]. In contrast, the network approach focuses on *functional connectivity*, the relationship between the neural activity of distinct brain regions [46]. While this connectivity can also be measured in a task-based experiment, our focus will be on resting-state studies in which subjects are told to relax and not think of anything in particular [153]. To appreciate the implications and limitations of functional connectivity studies, it is crucial to understand how it is measured. In this section, we briefly review the physics and biology behind fMRI and MEG, as well as the different metrics used to characterise functional connectivity.

Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) measures brain activity by detecting changes in blood flow [63]. Neural activity in a given region triggers a haemodynamic response, in which blood flow to the region increases temporarily in order to satisfy the local energy requirement. Importantly, this process results in an increase in the ratio of oxygenated haemoglobin to deoxygenated haemoglobin. The fMRI scanner exploits

the different magnetic properties of oxygenated haemoglobin compared to deoxygenated haemoglobin to localise neural activity at a given point in time.

The scanner generates three-dimensional volumes of the brain over time. Each volume is made up of voxels, cuboids with dimensions usually in the range of 2-4mm. (The voxel is the 3D equivalent of the more commonly-known 2D image unit, the pixel.) In a typical fMRI scan, a normal sized brain consists of roughly 100,000-150,000 voxels. The time taken to generate a single volume, known as the sampling rate or repetition time (TR), typically lies between TR = 1s and TR = 3s.

Given their large number, it is common to cluster voxels together into regions of interest (ROIs) [111]. This procedure, known as parcellation, serves to both reduce the dimensionality of the data and improve its interpretability. Many diverse parcellation schemes exist, varying both in the number of ROIs and in how these ROIs are defined [150, 66, 37]. When applied to an fMRI scan, the result of each of these schemes is a representative time series of neural activity for each ROI. Mathematically, this can be described as a multivariate time series $\{\mathbf{X}_t\}_{t=1}^{T}$ where T is the number of samples and each component $\{X_t^r\}_{t=1}^{T}$ corresponds to a ROI r. One can then investigate the brain's functional connectivity structure by studying the statistical interdependence of these component time series.

fMRI provides an indirect measure of neural activity (through blood flow) that is susceptible to various sources of noise such as head motion [64], scanner drift [136], and cardiovascular effects [34]. This is particularly troublesome in the context of group studies or heterogeneous samples; different groups or subjects may have differing haemodynamic properties or levels of noise. For example, the shape of the haemodynamic response function (HRF), which describes the blood flow following neural activity, is known to vary with age [73]. While extensive preprocessing pipelines serve to reduce the effects of these confounds, they cannot be mitigated completely. It is important to keep this fact in mind when comparing functional connectivity across groups as differences may arise due to variations in such confounds rather than true differences in the neural co-activation structure.

Magnetoencephalography

Magnetoencephalography (MEG) is another functional neuroimaging technique that measures the magnetic fields induced by neural activity [65]. Modern MEG scanners consist of around 300 sensors that detect the magnetic field at carefully selected locations around

the head. In contrast with fMRI, this provides a more direct measure of neural activity rather than a surrogate such as blood flow. On the other hand, the MEG scanner only provides recordings at the sensor locations, not at the true electrical sources inside the head that generate the magnetic fields outside the head.

The problem of transforming the data from *sensor space* to *source space* is known as an inverse problem. The challenge is to reconstruct the electrical activity from regions within the brain (source location) from the magnetic field detected at the sensor locations, where the number of regions is greater than the number of sensors. One popular approach for this is beamforming [157]. Beamforming is a spatial filtering technique that estimates the source activity at locations on a regular three-dimensional grid within the brain. The main assumption underlying beamforming is that the activity at distinct sources is uncorrelated. This somewhat limits the spatial resolution of MEG to several millimetres as sources close to each other are likely to exhibit correlated neural activity. As with fMRI, one can then perform a parcellation step to obtain an estimate of neural activity for ROIs from the beamformed data.

The temporal resolution of MEG, however, is typically on the order of milliseconds, making it capable of capturing extremely rapid oscillatory signals of brain activity [13]. Given the higher temporal resolution, it is typical to analyse MEG in different frequency bands: θ (4Hz to 7Hz), α (8Hz to 13Hz), β (13Hz to 30Hz), low γ (30Hz to 50Hz), and high γ (50Hz to 100Hz). Different frequency bands have been found to correspond to different neuronal processes [94].

Connectivity measures

In both the cases of fMRI and MEG, time series of measured neural activity for a number of ROIs are obtained. To analyse the functional connectivity structure of the ROIs requires a measure of co-activation between these time series. While there exists a plethora of different functional connectivity metrics [159, 17], each with their own strengths and weaknesses, we focus on Pearson correlation, which can be applied to both the amplitude of fMRI signal and the amplitude envelope of oscillatory activity in different frequency bands for MEG [71].

Pearson correlation is the most common measure of functional connectivity in fMRI [120]. Given two time series $\{X_t^r\}_{t=1}^T$ and $\{X_t^s\}_{t=1}^T$ corresponding to ROIs r and s, the

Pearson correlation between the two ROIs is given by

$$\rho_{r,s} = \frac{\sum_{t=1}^{T}(X_t^r - \bar{X}^r)(X_t^s - \bar{X}^s)}{\sqrt{\sum_{t=1}^{T}(X_t^r - \bar{X}^r)^2 \sum_{t=1}^{T}(X_t^s - \bar{X}^s)^2}} \tag{1.1}$$

where $\bar{X}^r = \frac{1}{T}\sum_{t=1}^{T} X_t^r$.

Amplitude envelope correlation measures the Pearson correlation between the logarithm of the amplitude envelopes of two signals. Let $X_r(t)$ be the continuous-time signal associated with ROI r. The corresponding analytic signal is given by

$$h_r(t) = X_r(t) + i\mathcal{H}[X_r(t)] \tag{1.2}$$

where $\mathcal{H}[X_r(t)]$ is the Hilbert transform of $X_r(t)$. The amplitude envelope of $X_r(t)$ is given by the magnitude of the analytic signal $|h_r(t)|$. From existing resting-state MEG functional connectivity metrics, amplitude envelope correlation was found to be the most reliable [31].

Time-varying functional connectivity

The measures described above are typically calculated for the entire duration of the scan. Recent evidence, however, suggests that, even in resting-state, these functional connections change over the course of a scan [7, 26, 80]. Moreover, dynamic functional connectivity (dFC) approaches, which measure such changes in connections, have been used to identify biomarkers for schizophrenia [122] and Alzheimer's disease [78].

The majority of dFC studies apply a sliding-window analysis to the ROI time series. This involves creating w equally spaced (tapered) windows of the same size which cover the length of the time series. For each window, connectivity between ROIs (typically Pearson correlation) is calculated for the windowed time series consisting only of the time points within that window. The windows are typically overlapping and thus able to detect faster fluctuations compared to disjoint windows, at the cost of introducing dependency between connectivity estimates in nearby windows. This yields a time series of length w consisting of connectivity measures (correlation) for each pair of ROIs.

Methods to analyse the resulting correlation time series range from the relatively simple (e.g. characterising dFC as the standard deviation or power of the correlation time series [42]) to the more sophisticated (e.g. identifying recurring connectivity patterns

using k-means clustering [7]). Most approaches, however, do not account for heterogeneity between individuals. Heterogeneity can arise in a variety of ways, including vascular differences or degrees of head motion, which may have a range of effects on observed functional connectivity dynamics. While the effect of such heterogeneity on static connectivity has been studied previously [95, 11, 114], the impact on dFC measures is less explored.

1.2 Networks in neuroimaging

Network construction

The measures of functional connectivity defined in the previous section quantify pairwise interactions between any two ROIs. The extension to whole-brain functional connectivity is straightforward: compute the pairwise metric between each pair of ROIs. This yields a symmetric connectivity matrix $C \in \mathbb{R}^{R\times R}$ where R is the number of ROIs. The connectivity matrix can be associated with a weighted network with nodes corresponding to each column or row, and edge weights given by the corresponding entries in the matrix. Note that, with probability one, all of the entries in C are non-zero and so the associated network is fully connected. This network, however, is likely to include a significant proportion of spurious connections due to, for example, the presence of measurement noise [1].

This problem is addressed by applying a threshold to the connectivity matrix to set all weak or negative connections to zero. To gain insight into the topological properties of the brain's connectivity structure, it is typical to then binarise the resulting network by setting all remaining non-zero connections to one. Specifically, the binarisation procedure is performed by applying a threshold c_{thresh} to the connectivity matrix C such that a connection exists between ROIs r and s if and only if $C_{rs} \geq c_{thresh}$:

$$A_{rs} = \begin{cases} 1 & \text{if } C_{rs} \geq c_{thresh} \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

The adjacency matrix $A \in \{0,1\}^{R\times R}$ defines a graph: $A_{rs} = 1$ if an edge (connection) exists between nodes (ROIs) r and s, and $A_{rs} = 0$ otherwise. An example of a binarised functional connectivity network is shown in Figure 1.1.

In constructing a network from functional brain imaging data, there are thus a num-

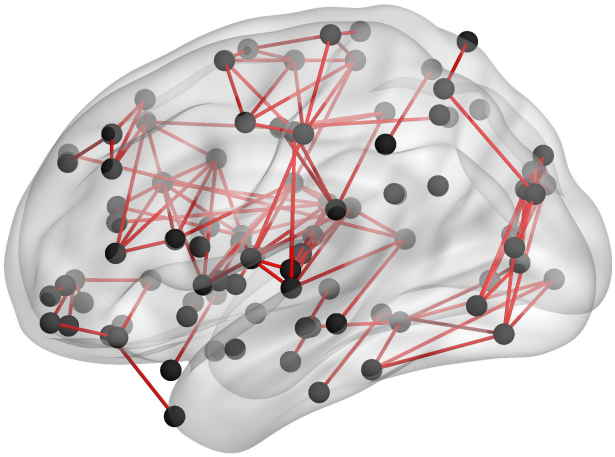


Figure 1.1: An example of a binarised functional connectivity network. This image was created using BrainNet Viewer [164].

ber of choices to be made, each of which will have an impact on the subsequent analysis. First, to reduce the dimensionality of the data, the brain is parcellated into distinct brain regions and a representative time course constructed for each region. The regions define the nodes of the network and different parcellations yield networks of different sizes. Second, weighted edges are defined via a pairwise functional connectivity metric, typically Pearson correlation, between regional time courses. Thirdly, these edges are thresholded and binarised to yield a sparse, unweighted network. The choice of threshold also has important implications for the resulting network structure [156, 60]. While there are a variety of thresholding schemes, the most common approach is to select c_{thresh} to yield a network of a given density, typically between 5% and 50% [44].

There is no single correct way to construct a network from functional brain imaging data; each of the choices outlined above should be motivated by the scientific question at hand. Nevertheless, the framework presented in this thesis is applicable to any set of networks, regardless of how they are constructed.

Network modelling

The application of graph theory to brain connectivity, both structural and functional, has become increasingly popular in recent years [22]. Early graph-theoretical studies focused on comparisons to random networks via the computation of various network metrics. For

example, functional brain networks as measured by both fMRI and MEG were found to have a shorter average path length and a higher clustering coefficient than expected in a random network [2, 139]. Network metrics were then employed to quantify differences between groups of individuals, for example young people versus old people [98] and patients with Alzheimer's disease versus healthy subjects [141].

This approach suffers from a serious limitation: network metrics are influenced by the overall density of the network [62]. As a result, it is difficult to disentangle differences in more complex topological properties such as clustering from differences in the network density. This is particularly important in the context of ageing since mean functional connectivity is known to decrease with age [54]. Note that it is not trivial to simply control for network density when comparing network metrics due to their highly non-linear relationship.

One possible means to address this issue is to use exponential random graph models (ERGMs). An exponential random graph model is a set of parametric statistical distributions on network data (see [118] for a review). The aim of the model is to characterise the distribution of a network in terms of a set of *summary statistics*, or network metrics. This is especially useful in the context of neuroimaging: a better understanding of how (local) topological features give rise to the brain's global network structure could yield crucial insights into the processes that underlie cognitive function. Moreover, by including network density along with other metrics of interest in the model, ERGMs provide a way to quantify their relative influence on the overall network structure while accounting for the density.

ERGMs have been applied successfully to characterise both functional connectivity [132, 131, 104] and structural connectivity [134]. To date, there have been two proposed approaches for using ERGMs in group studies. The first approach constructs a single group network by keeping edges that are present in a minimum of the individual networks, and then fits an ERGM to the group network [134]. The main drawback of this approach lies in the averaging over information present in the individual networks. In contrast, the second approach fits an ERGM to each individual network and then takes the mean or median of the fitted parameters to represent the group-level connectivity structure [131, 104]. While this is preferable to the first approach, taking a simple mean or median seems oversimplistic. The main contribution of this thesis is to develop a multilevel framework for ERGMs, thus allowing information to be pooled across networks and providing a mechanism to infer group-level differences in connectivity structure.

1.3 The Cam-CAN study

The motivating application for this thesis derives from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) research project [127], which aims to improve understanding of the effect of healthy ageing on cognitive and brain function. The Cam-CAN dataset consists of a range of cognitive tests and functional neuroimaging experiments for approximately 650 healthy individuals aged 18-87 (roughly 100 per age decade, though with fewer in the youngest decade). Our focus will be on the youngest decade (18-27) and the oldest decade (78-87); the aim will be to compare the functional connectivity structure between these two groups.

1.4 Outline of thesis

The core of the thesis is contained in Chapters 2-5. Chapter 2 introduces the main methodological development: a multilevel framework for populations of networks based on exponential random graph models. Chapter 3 applies this framework to resting-state functional connectivity networks constructed from data from the Cam-CAN study. Using this dataset, we illustrate how our method can be used to detect differences in functional connectivity between a group of young individuals and a group of old individuals. In Chapter 4, we shift our focus to dynamic functional connectivity and develop a generic simulation framework for fMRI data. We then use simulations to investigate the effect of various sources of heterogeneity on estimates of dFC. Chapter 5 extends the multilevel framework for use with time-varying connectivity networks. Using MEG data from the Cam-CAN study, we apply the extended method to detect differences in time-varying connectivity between a young group and an old group. Finally, Chapter 6 discusses some of the challenges and limitations of the work presented in the thesis, and identifies some avenues for future work.

Chapter 2

A multilevel framework for populations of networks using exponential random graph models

Summary

Most statistical network models were originally proposed to describe a single underlying relational structure such as friendships between individuals or hyperlinks between web pages. As a result, the development of these models has largely been restricted to the single network case. Most neuroimaging datasets, however, consist of multiple networks across several individuals. For such studies, a typical goal is to infer a common connectivity structure across individuals. While one could in principle fit a single network model to each individual separately, it is not always straightforward to combine these individual results into a single group result. Using Bayesian exponential random graph models, we propose a multilevel framework for populations of networks. By pooling information across the individual networks, this framework provides a principled approach to characterise the relational structure for an entire population.

This chapter consists of the following: Section 2.1 presents the exponential random graph model (ERGM); Section 2.2 discusses Bayesian inference for ERGMs; Section 2.3 introduces both the multilevel framework extending the Bayesian ERGM to populations of networks and the model fitting procedure; Section 2.4 outlines how we perform model assessment within the framework; and Section 2.5 describes improvements to the fitting procedure in the name of computational efficiency.

2.1 Exponential random graph models

An exponential random graph model is a set of parametric statistical distributions on network data (see [118] for a review). The aim of the model is to characterise the distribution of a network in terms of a set of *summary statistics*. These summary statistics are typically comprised of topological features of the network, such as the number of edges and subgraph counts. The summary statistics enter the likelihood via a weighted sum; the weights are (unknown) model parameters that quantify the relative influence of the corresponding summary statistic on the overall network structure and must be inferred from the data. ERGMs are thus a flexible way in which to describe the global network structure as a function of network summary statistics.

Notation

We first introduce some notation and terminology. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a random network consisting of a *fixed* set $\mathcal{V}$ of nodes and a *random* set $\mathcal{E}$ of undirected edges. Write $V = |\mathcal{V}|$ for the (fixed) number of nodes of $\mathcal{G}$. The sample space of $\mathcal{E}$ is $\Omega = \{0,1\}^{\frac{V(V-1)}{2}}$; there can either be an edge (1) or not an edge (0) between the $\frac{V(V-1)}{2}$ pairs of nodes in $\mathcal{V}$. Write $E = |\mathcal{E}|$ for the number of edges in $\mathcal{G}$. Note that E itself is a random variable that can take values between 0 (an empty network) and $\frac{V(V-1)}{2}$ (a complete network).

We will exclusively work with the adjacency matrix $\mathbf{Y}$ associated with the network $\mathcal{G}$. (We will use the term network to refer to both $\mathcal{G}$ and $\mathbf{Y}$.) The adjacency matrix $\mathbf{Y}$ is a symmetric random matrix taking values in $\mathcal{Y} = \{0,1\}^{V \times V}$. Specifically, $Y_{ij} = Y_{ji} = 1$ if there is an edge $e \in \mathcal{E}$ between nodes $i \in \mathcal{V}$ and $j \in \mathcal{V}$. Denote $\mathbf{y}$ to be an instantiation, or outcome, of the random adjacency matrix $\mathbf{Y}$ and write $\mathbb{P}(\mathbf{Y} = \mathbf{y}) := \pi(\mathbf{y})$ for the probability that $\mathbf{Y}$ takes the value $\mathbf{y}$. Let $\mathcal{Y}$ be the range of $\mathbf{Y}$, i.e. the set of all possible outcomes. Let $s(\mathbf{Y})$ denote a vector of p summary statistics of $\mathbf{Y}$, such that each component is a function $s_i : \mathcal{Y} \to \mathbb{R}$.

We reserve non-italicised boldface type for matrices (such as $\mathbf{Y}$ and $\mathbf{\Sigma}$) and generally use non-boldface type for scalars and vectors.

Summary statistics

In the context of ERGMs, a summary statistic is simply a network metric, i.e. a real-valued function of some configuration of edges in the network. Simple examples include

Summary statistic	**Description**
edges	Number of edges
k-cycle	Number of k-cycles
k-star	Number of k-stars
k-degree	Number of nodes with degree k
geometrically weighted degree distribution (GWDEG)	Weighted degree distribution with weight for k-degree given by $(1 + \exp(-\tau))^k$ where τ is a fixed decay parameter
k dyadwise shared partners	Number of node pairs having exactly k shared partners
geometrically weighted dyadwise shared partner distribution (GWDSP)	Weighted distribution of dyadwise shared partners with weight for k edgewise shared partners given by $(1 + \exp(-\tau))^k$ where τ is a fixed decay parameter
k edgewise shared partners	Number of *connected* node pairs having exactly k shared partners
geometrically weighted edgewise shared partner distribution (GWESP)	Weighted distribution of edgewise shared partners with weight for k edgewise shared partners given by $(1 + \exp(-\tau))^k$ where τ is a fixed decay parameter
k non-edgewise shared partners	Number of *unconnected* node pairs having exactly k shared partners
geometrically weighted non-edgewise shared partner distribution (GWNSP)	Weighted distribution of non-edgewise shared partners with weight for k non-edgewise shared partners given by $(1 + \exp(-\tau))^k$ where τ is a fixed decay parameter

Table 2.1: A selection of summary statistics commonly used in ERGMs.

the number of edges or triangles. (Some other common choices can be found in Table 2.1.) The summary statistics included in a given ERGM represent those configurations expected to appear more frequently or less frequently than in a random graph. In other words, the choice of summary statistics is a modelling decision; it reflects our belief of how the global network structure may be summarised and is driven by the context of the network. The flexibility of ERGMs derives from the range and number of possible summary statistics that can be used [117]. Note that each distinct set of summary statistics (up to rescaling) defines a different ERGM.

Definition

We are now equipped to define an exponential random graph model. Let $\mathbf{Y}$ be a network, and let $s(\mathbf{Y})$ be a vector-valued function of p summary statistics defined on $\mathcal{Y}$. The probability mass function of $\mathbf{Y}$ under the corresponding ERGM is given by $\pi(\mathbf{y}|\theta)$ where

$$\pi(\mathbf{y}|\theta) = \frac{\exp\left\{\theta^T s(\mathbf{y})\right\}}{Z(\theta)}. \tag{2.1}$$

Here, $\theta \in \Theta \subseteq \mathbb{R}^p$ is a vector of p model parameters that must be estimated from the data and $Z(\theta) = \sum_{\mathbf{y}' \in \mathcal{Y}} \exp\left\{\theta^T s(\mathbf{y}')\right\}$ is the normalising constant ensuring the probability mass function sums to one. Given data, that is, an observation $\mathbf{y}$ of $\mathbf{Y}$, the goal is to infer which values of θ best correspond to the data under this distribution.

Remarks

Exponential random graph models are a flexible family of distributions on networks that aim to characterise the global network structure in terms of a relatively small number of summary statistics on the network. This is especially useful in the context of neuroimaging: a better understanding of how (local) topological features give rise to the brain's global network structure could yield crucial insights into the processes that underlie cognitive function. One particular challenge in network neuroscience arises from the fact that network metrics are influenced by the overall density of the network [62]. In the context of ageing, for example, mean functional connectivity is known to decrease with age [54], making it difficult to disentangle differences in more complex topological properties such as clustering from differences in the network density. By including network density along with other metrics of interest in the model, ERGMs provide a way to quantify their relative influence on the overall network structure while accounting for the density.

The flexibility of the family stems from the range and number of summary statistics that can be included in the model. Given a choice of summary statistics, the model parameters θ quantify the relative influence of the corresponding statistic on the overall structure of the network. To see this, consider the log-odds of an edge Y_{ij}, conditional on observing the rest of the network. Denote $\mathbf{Y}_{ij}^c = \mathbf{Y} \setminus Y_{ij}$ to be the set of variables in the random adjacency matrix $\mathbf{Y}$ excluding that of the random variable Y_{ij} associated with the single node pair (i, j). Denote $\mathbf{y}_{ij}^c$ to be the corresponding observations of $\mathbf{Y}_{ij}^c$. Then,

$$\begin{aligned}\log\left(\frac{\mathbb{P}(Y_{ij}=1|\mathbf{Y}^c_{ij}=\mathbf{y}^c_{ij})}{\mathbb{P}(Y_{ij}=0|\mathbf{Y}^c_{ij}=\mathbf{y}^c_{ij})}\right) &= \log\left(\frac{\mathbb{P}(Y_{ij}=1,\mathbf{Y}^c_{ij}=\mathbf{y}^c_{ij})}{\mathbb{P}(Y_{ij}=0,\mathbf{Y}^c_{ij}=\mathbf{y}^c_{ij})}\right)\\ &= \theta^T\{s(\mathbf{y}^+_{ij})-s(\mathbf{y}^-_{ij})\}\end{aligned} \tag{2.2}$$

where $\mathbf{y}^+_{ij}$ is the network with $Y_{ij}=1$ and $\mathbf{Y}^c_{ij}=\mathbf{y}^c_{ij}$, and similarly $\mathbf{y}^-_{ij}$ is the same network except $Y_{ij}=0$. Thus, each component of θ can be interpreted as the difference in conditional log-odds per unit change in the corresponding summary statistic between the networks $\mathbf{y}^+_{ij}$ and $\mathbf{y}^-_{ij}$. All else being equal, a larger value of the θ component places more weight on networks $\mathbf{y}$ with a larger value of the corresponding summary statistic.

The main barrier to inference is the normalising constant $Z(\theta)$. This is the sum of the numerator over all possible networks in $\mathcal{Y}$ and so, even for moderately sized networks, is typically intractable. For example, a network with just 10 nodes has $2^{\frac{10(10-1)}{2}} = 35,184,372,088,832$ possible network configurations. As a result, computing the sum $Z(\theta)=\sum_{\mathbf{y}'\in\mathcal{Y}}\exp\left\{\theta^T s(\mathbf{y}')\right\}$ over all these configurations is infeasible. We will thus refer to $Z(\theta)$ as an *intractable normalising constant*, or *INC*, because we cannot compute it in sensible time.

Although ERGMs are a powerful and flexible class of distributions, they have a number of limitations. For example, ERGMs have been found to suffer from degeneracy issues [137, 67]. This refers to the model estimation problems that result from certain model specifications. Specifically, under such specifications, the model places disproportionate mass on empty or fully connected networks. Since inference methods for ERGMs rely on MCMC simulation from the model, the result is that the chain moves to and stays at an empty or full network, irrespective of the observed network. Degeneracy occurs as a result of model misspecification and may be overcome by using summary statistics that are less prone to degeneracy issues [138]. Another limitation is that ERGM coefficients are not necessarily comparable across networks of different sizes [87]. Thus, our proposed framework is restricted to the case where each of the networks have the same number of nodes. This is not a major problem in the context of a neuroimaging study because the same parcellation is typically applied to each brain scan, resulting in the same number of nodes.

2.2 Bayesian inference for ERGMs

In what follows, we will work in the Bayesian paradigm, treating the model parameters θ as random variables. The Bayesian formulation of ERGMs was first suggested in [82] and then expanded upon in [24]. Through the machinery of Bayesian hierarchical modelling, we shall see that this provides a natural framework for *populations* of networks, as typically found in neuroimaging datasets.

To fully specify a Bayesian ERGM, we need only augment the definition in (2.1) with a prior distribution $\pi(\theta)$ for the model parameters. Given an observation $\mathbf{y}$ of the network, we can then perform inference by analysing the posterior distribution $\pi(\theta|\mathbf{y})$. Through Bayes' rule, we may write the posterior as

$$\begin{aligned}\pi(\theta|\mathbf{y}) &= \frac{\pi(\mathbf{y}|\theta)\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})} \\ &= \frac{\exp\left\{\theta^T s(\mathbf{y})\right\}\pi(\theta)}{Z(\theta)\pi(\mathbf{y})}\end{aligned} \tag{2.3}$$

where $\pi(\mathbf{y}) = \int_\Theta \pi(\mathbf{y}|\theta)\pi(\theta)d\theta$ is the model evidence. The posterior distribution is generally not available in a closed-form expression. This is due to two properties of the posterior: first, the (standard) intractability of the model evidence, and second, the intractability of the likelihood via the INC $Z(\theta)$. Posterior distributions with these two sources of intractability are referred to as *doubly-intractable*.

Although the distribution is not available in closed-form, we can nevertheless use numerical methods to study the posterior. Our aim is to generate samples from the posterior distribution and use these to calculate numerical approximations to certain integrals, such as the posterior mean or credible intervals. To do so, we will use Markov chain Monte Carlo (MCMC) algorithms. The key idea behind MCMC is to construct a Markov chain that has the posterior as its stationary distribution. Then, after a sufficient number of steps, samples from this chain should have the desired distribution.

Standard MCMC approaches fail

Unfortunately, when the target distribution is doubly-intractable, standard MCMC approaches tend to fail. To see why, consider the Metropolis algorithm [97] with symmetric proposal function $h(\theta'|\theta)$, outlined in Algorithm 2.1. It is straightforward to show that the stationary distribution of the Markov chain produced by this algorithm is $\pi(\theta|\mathbf{y})$.

Algorithm 2.1 Metropolis algorithm for a Bayesian ERGM

Require: number of MCMC iterations K, initial value θ_0

for $k = 1, \ldots, K$ **do**

- draw θ' from $h(\cdot|\theta_{k-1})$
- calculate acceptance probability

$$AR(\theta', \theta_{k-1}; \mathbf{y}) = \frac{\pi(\theta'|\mathbf{y})}{\pi(\theta_{k-1}|\mathbf{y})}$$

- set $\theta_k = \theta'$ with probability $\min(1, AR(\theta', \theta_{k-1}))$
- else, set $\theta_k = \theta_{k-1}$

end for

The acceptance ratio $AR(\theta', \theta_{k-1}; \mathbf{y})$ in Algorithm 2.1 can be rewritten as

$$\begin{aligned} AR(\theta', \theta_{k-1}; \mathbf{y}) &= \frac{\pi(\theta'|\mathbf{y})}{\pi(\theta_{k-1}|\mathbf{y})} \\ &= \frac{\exp\left\{\theta'^T s(\mathbf{y})\right\} \pi(\theta')}{\exp\left\{\theta_{k-1}^T s(\mathbf{y})\right\} \pi(\theta_{k-1})} \cdot \frac{Z(\theta_{k-1})}{Z(\theta')}. \end{aligned} \tag{2.4}$$

As intended, the model evidence $\pi(\mathbf{y})$ cancels out. However, the acceptance ratio still contains the intractable normalising constant ratio (INCR) $Z(\theta_{k-1})/Z(\theta')$. Herein lies the main obstacle of carrying out MCMC methods for Bayesian inference in the presence of an intractable likelihood: the transition probabilities may not be computable. As a result, it is not possible to produce sequences of samples from the desired Markov chain.

The exchange algorithm

To circumvent this obstacle, the exchange algorithm [102] can be adapted for Bayesian ERGMs [24]. This is outlined in Algorithm 2.2. The exchange algorithm is an MCMC scheme designed to address the INCR in the acceptance probability (2.4). This is achieved by introducing an auxiliary variable $\mathbf{y}' \sim \pi(\cdot|\theta')$, i.e. a network drawn from the same exponential random graph model with parameter θ'.

The algorithm works by targeting an augmented posterior

$$\pi(\theta, \theta', \mathbf{y}'|\mathbf{y}) \propto \pi(\theta|\mathbf{y}) h(\theta'|\theta) \pi(\mathbf{y}'|\theta') \tag{2.5}$$

Algorithm 2.2 The exchange algorithm for a Bayesian ERGM [24]

Require: number of MCMC iterations K, initial value θ_0

for $k = 1, \ldots, K$ **do**

- draw $\theta' \sim h(\cdot|\theta_{k-1})$
- draw $\mathbf{y}' \sim \pi(\cdot|\theta')$
- set $\theta_k = \theta'$ with probability $\min\left(1, AR(\theta', \theta_{k-1}, \mathbf{y}, \mathbf{y}')\right)$ ▷ See Eq. (2.6)
- else, set $\theta_k = \theta_{k-1}$

end for

where $\pi(\theta|\mathbf{y})$ is the original (target) posterior, $h(\theta'|\theta)$ is an arbitrary normalisable proposal function, and $\pi(\mathbf{y}'|\theta)$ is the likelihood of the auxiliary variable. For simplicity, we assume $h(\theta'|\theta)$ to be symmetric. Since each of the three terms on the right-hand side of Eq. (2.5) can be normalised, the left-hand side is well-defined as a probability distribution.

The algorithm proceeds as follows. At each iteration, first perform a Gibbs' update of $(\theta', \mathbf{y}')$ by drawing $\theta' \sim h(\cdot|\theta)$ followed by $\mathbf{y}' \sim \pi(\cdot|\theta')$. Next, *exchange* θ and θ' with probability $\min(1, AR(\theta', \theta, \mathbf{y}, \mathbf{y}'))$, where

$$\begin{aligned} AR(\theta', \theta, \mathbf{y}, \mathbf{y}') &= \frac{\pi(\theta'|\mathbf{y})}{\pi(\theta|\mathbf{y})} \cdot \frac{\pi(\mathbf{y}'|\theta)}{\pi(\mathbf{y}'|\theta')} \\ &= \frac{\exp\left\{\theta'^T s(\mathbf{y})\right\} \pi(\theta')}{\exp\left\{\theta^T s(\mathbf{y})\right\} \pi(\theta)} \frac{Z(\theta)}{Z(\theta')} \cdot \frac{\exp\left\{\theta^T s(\mathbf{y}')\right\}}{\exp\left\{\theta'^T s(\mathbf{y}')\right\}} \frac{Z(\theta')}{Z(\theta)} \\ &= \exp\left\{[\theta' - \theta]^T [s(\mathbf{y}) - s(\mathbf{y}')]\right\} \frac{\pi(\theta')}{\pi(\theta)} \end{aligned} \tag{2.6}$$

Crucially, the INCRs cancel out, and so this acceptance ratio can indeed be evaluated. The stationary distribution of the Markov chain constructed through this scheme is $\pi(\theta, \theta', \mathbf{y}'|\mathbf{y})$ [102]. Thus, by marginalising out θ' and $\mathbf{y}'$, the algorithm yields samples from the desired posterior, namely $\pi(\theta|\mathbf{y})$.

Remarks

At each iteration, the exchange algorithm requires a sample $\mathbf{y}'$ from the ERGM $\pi(\cdot|\theta')$ in order to compute the acceptance ratio. This is in fact the main computational bottleneck in running the algorithm. Simulating from ERGMs is challenging and existing sampling schemes are computationally intensive. The most common approach is to use a Metropolis-Hastings algorithm [70, 74], as follows. At each iteration, propose a network

transition (by, for example, toggling an edge on or off) and accept the transition with a certain probability. This probability is chosen such that the algorithm produces a Markov chain of networks that has stationary distribution given by $\pi(\cdot|\theta')$. One can then take a network from the sequence, after a sufficient number of steps, as an approximate draw from $\pi(\cdot|\theta')$.

As a result, Algorithm 2.2 comprises a *nested MCMC*: for each proposal θ', we must perform an inner MCMC in order to generate an approximate sample from $\pi(\cdot|\theta')$. A theoretical justification of this approach was given by Everitt [43]: under certain conditions, despite using an approximate sample, the algorithm nevertheless targets an approximation to the correct posterior distribution. Further, this approximation improves as the number of iterations of the inner MCMC increases. Of course, the efficient convergence of the inner MCMC is crucial for the efficiency of the overall algorithm. The convergence of MCMC algorithms for ERGMs has been studied in some detail. In particular, unusual convergence properties have been observed for some combinations of summary statistics [137], and under certain regimes, the mixing time is exponential in the number of nodes for MCMC algorithms with local transitions [18]. Therefore, the application of Bayesian ERGMs has been restricted to networks with around 100 nodes or less.

2.3 A framework for populations of networks

The Bayesian exponential random graph model described above provides a flexible family of distributions for a *single* network. Our goal is to extend this to a model for a *population* of networks. Our proposed approach is simple: represent each network as a separate ERGM within a Bayesian multilevel (or hierarchical) model. By pooling information across individual networks, this approach allows us to characterise the distribution of the whole population.

Notation

Before fully describing the framework, we must first introduce some more notation. Let $\boldsymbol{Y} = (\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(n)})$ be a set of n networks. Identify each network $\mathbf{Y}^{(i)}$ with its own *individual-level* (and vector-valued) ERGM parameter $\theta^{(i)}$. Write $\boldsymbol{\theta} = (\theta^{(1)}, \ldots, \theta^{(n)})$ for the set of individual-level parameters. We reserve italicised boldface type for sets of matrices (such as $\boldsymbol{Y}$) and sets of vectors (such as $\boldsymbol{\theta}$).

Description of framework

We begin by deriving the sampling distribution $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ for the population of networks $\boldsymbol{Y}$. We model each individual network $\mathbf{Y}^{(i)}$ as an exponential random graph with model parameter $\theta^{(i)}$. Importantly, each individual ERGM must consist of the same set of summary statistics $s(\cdot)$. The probability mass function of each network can then be written

$$\pi(\mathbf{y}^{(i)}|\theta^{(i)}) = \frac{\exp\left\{\theta^{(i)T} s(\mathbf{y}^{(i)})\right\}}{Z(\theta^{(i)})}, \quad i = 1, \ldots, n. \tag{2.7}$$

This specifies the data-generating process for each individual network. To obtain a joint distribution for the set of networks, we assume that, conditional on their respective individual-level parameters, the $\mathbf{Y}^{(i)}$ are independent. Thus, the sampling distribution for the set of networks $\boldsymbol{Y}$ is simply the product of the individual probability mass functions:

$$\begin{aligned} \pi(\boldsymbol{y}|\boldsymbol{\theta}) &= \prod_{i=1}^{n} \pi(\mathbf{y}^{(i)}|\theta^{(i)}) \\ &= \frac{\exp\left\{\sum_{i=1}^{n} \theta^{(i)T} s(\mathbf{y}^{(i)})\right\}}{\prod_{i=1}^{n} Z(\theta^{(i)})}. \end{aligned} \tag{2.8}$$

To complete the model, we need to specify the prior distribution of the individual-level ERGM parameters $\theta^{(1)}, \ldots, \theta^{(n)}$. To this end, we posit a 2-stage multilevel model: assume the $\theta^{(1)}, \ldots, \theta^{(n)}$ are drawn from a common population-level distribution with hyperparameter ϕ, which itself is a random variable. We write

$$\theta^{(i)} \sim \pi(\cdot|\phi), \quad i = 1, \ldots, n \tag{2.9}$$

for the population-level distribution. Assuming that, conditional on the hyperparameter ϕ, the $\theta^{(i)}$ are independent, we have

$$\pi(\boldsymbol{\theta}|\phi) = \prod_{i=1}^{n} \pi(\theta^{(i)}|\phi). \tag{2.10}$$

Finally, write $\pi(\phi)$ for the (hyper)prior distribution of ϕ. The joint distribution of $(\boldsymbol{Y}, \boldsymbol{\theta}, \phi)$ can be written as $\pi(\boldsymbol{y}, \boldsymbol{\theta}, \phi) = \pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\phi)\pi(\phi)$. See Figure 2.1 for a diagrammatic representation of the full model.

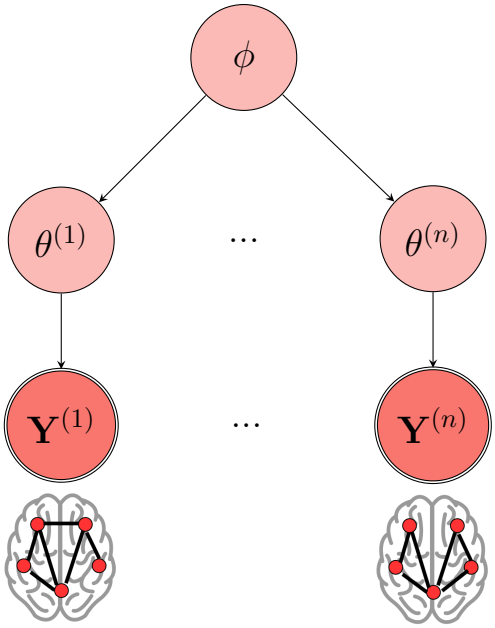


Figure 2.1: A diagrammatic representation of the hierarchical framework. Each network $\mathbf{Y}^{(i)}$ is modelled as an exponential random graph with individual-level parameter $\theta^{(i)}$. In turn, each θ_i is assumed to come from a common population-level distribution with hyperparameter ϕ.

Applying Bayes' theorem, we obtain the (joint) posterior for $(\phi, \boldsymbol{\theta})$:

$$\begin{aligned}\pi(\boldsymbol{\theta}, \phi|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}, \phi)\pi(\boldsymbol{\theta}, \phi)}{\pi(\boldsymbol{y})} \\ &= \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\phi)\pi(\phi)}{\pi(\boldsymbol{y})} \\ &= \frac{\pi(\phi)\exp\left\{\sum_{i=1}^{n}\theta^{(i)T}s(\mathbf{y}^{(i)})\right\}\prod_{i=1}^{n}\pi(\theta^{(i)}|\phi)}{\pi(\boldsymbol{y})\prod_{i=1}^{n}Z(\theta^{(i)})}.\end{aligned} \tag{2.11}$$

where $\pi(\boldsymbol{y}) = \int \pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the model evidence. As with the Bayesian ERGM for a single network, this posterior is doubly-intractable: both the model evidence $\pi(\boldsymbol{y})$ and the likelihood $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ (via the INCs $Z(\theta^{(1)}, \ldots, Z(\theta^{(n)}))$ are analytically intractable.

Prior specification

The choice of prior distributions for Bayesian ERGMs has yet to be studied in any great detail. The appropriate setting of priors is a challenging task due to the typically high levels of dependence between parameters [83]. Studies thus far have generally assumed

(flat) multivariate normal prior distribution on the model parameters [24, 134, 146]. For simplicity, we also assume multivariate normal priors, though alternative specifications fully warrant further investigation. Specifically, we assume

$$\theta^{(i)} \sim \mathcal{N}(\mu, \mathbf{\Sigma}_\theta), \quad i = 1, \ldots, n. \tag{2.12}$$

Thus, in the above notation, we have $\phi = (\mu, \mathbf{\Sigma}_\theta)$.

For the hyperprior distribution of $\phi = (\mu, \mathbf{\Sigma}_\theta)$, we assume conditionally conjugate priors on μ and $\mathbf{\Sigma}_\theta$, placing a multivariate normal prior on μ and an inverse-Wishart prior on $\mathbf{\Psi}$:

$$\begin{aligned} \mu &\sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_\mu) \\ \mathbf{\Sigma}_\theta &\sim \mathcal{W}^{-1}(\mathbf{\Psi}_\theta, \nu_\theta) \end{aligned} \tag{2.13}$$

where $\mu_0, \nu, \mathbf{\Psi}$ are as above, and $\mathbf{\Sigma}_\mu \in \mathbb{R}^{p \times p}$ is a positive definite matrix. These choices of hyperpriors are again motivated by simplicity as well as computational convenience.

An alternative, fully conjugate, hyperprior is the normal-inverse-Wishart distribution on $(\mu, \mathbf{\Sigma}_\theta)$

$$(\mu, \mathbf{\Sigma}_\theta) \sim NIW(\mu_0, \lambda, \mathbf{\Psi}_\theta, \nu_\theta) \tag{2.14}$$

where $\mu_0 \in \mathbb{R}^p$, $\lambda > 0$, $\nu > p-1$, and $\mathbf{\Psi} \in \mathbb{R}^{p \times p}$ is a positive definite matrix. In the conditionally conjugate case, μ and $\mathbf{\Sigma}_\theta$ are a prior independent, whereas in the fully conjugate case, we have

$$\mu | \mathbf{\Sigma}_\theta \sim N\left(\mu_0, \frac{\mathbf{\Sigma}_\theta}{\lambda}\right). \tag{2.15}$$

In other words, the fully conjugate case may be more appropriate if the prior belief is that the covariance of μ is the same, up to a scalar multiple, as the covariance of the $\theta^{(i)}$. In practice, initial investigations indicated no substantial differences between the conditionally conjugate and fully conjugate cases in the resulting posteriors. Since it is more flexible and easier to extend to further levels of hierarchy, we therefore opted for the conditionally conjugate hyperprior instead of the fully conjugate hyperprior.

Model fitting: the exchange-within-Gibbs algorithm

In order to perform inference through this model, we would like to generate samples from its posterior distribution $\pi(\boldsymbol{\theta}, \phi | \boldsymbol{y})$. This posterior is doubly-intractable so, as discussed in Section 2.2, standard MCMC schemes such as the Metropolis algorithm are not suitable.

Although the exchange algorithm can be readily used for a single Bayesian ERGM, it has not (to our knowledge) been adapted for use within a Bayesian multilevel model. We therefore propose a novel MCMC scheme, the *exchange-within-Gibbs algorithm*, for this purpose.

As the name suggests, the exchange-within-Gibbs algorithm combines the exchange algorithm with the Gibbs sampler [57] to produce samples from the desired posterior. Note that we can treat the unknown parameters of the model $(\boldsymbol{\theta}, \phi)$ as *components* of a single multi-dimensional parameter. The idea behind Gibbs sampling is to iteratively sample each component from its conditional distribution given the remaining components. Specifically, at the k^{th} iteration, sample

$$\phi_k \sim \pi(\cdot|\boldsymbol{\theta}_{k-1}; \boldsymbol{y}) \tag{2.16}$$

$$\theta_k^{(i)} \sim \pi(\cdot|\boldsymbol{\theta}_{k-1}^{(-i)}, \phi_k; \boldsymbol{y}), \quad i = 1, \ldots, n. \tag{2.17}$$

where $\boldsymbol{\theta}^{(-i)} = (\theta^{(1)}, \ldots, \theta^{(i-1)}, \theta^{(i+1)}, \ldots, \theta^{(n)})$. Note that, since ϕ is conditionally independent of the data $\boldsymbol{Y}$ given $\boldsymbol{\theta}$, and the $\theta^{(i)}$ are conditionally independent given ϕ, equations (2.16) and (2.17) simplify to

$$\phi_k \sim \pi(\cdot|\boldsymbol{\theta}_{k-1}) \tag{2.18}$$

$$\theta^{(i)} \sim \pi(\cdot|\phi_k; \mathbf{y}^{(i)}), \quad i = 1, \ldots, n \tag{2.19}$$

respectively. This procedure produces a Markov chain of samples $(\boldsymbol{\theta}_k, \phi_k)$ with the posterior as stationary distribution. We now elaborate on these two Gibbs updating steps.

The precise updating step for ϕ depends on the choice of prior and hyperprior distributions. In general, it should be feasible to perform a standard Metropolis-Hastings step in order to draw $\phi_k \sim \pi(\cdot|\boldsymbol{\theta}_{k-1})$. Moreover, given conjugate priors, it may be possible to sample ϕ_k directly. For example, in the case of a multivariate normal prior with conditionally conjugate hyperpriors on μ and $\boldsymbol{\Sigma}_\theta$, we can write down the conditional posterior distributions in closed-form as a multivariate normal distribution for μ and an inverse-Wishart distribution for $\boldsymbol{\Sigma}_\theta$.

The second step is to sample from the conditional posterior of the individual-level parameters $\theta^{(i)}$ for $i = 1, \ldots, n$. Since this conditional posterior is doubly-intractable, we cannot use a standard Metropolis-Hastings update. We can, however, apply the exchange algorithm. Specifically, at the k^{th} iteration and for the i^{th} network, draw a proposal θ' from a (symmetric) proposal function $h(\cdot|\theta_{k-1}^{(i)})$. Next, sample an exponential random

graph $\mathbf{y}'$ with parameter θ'. With probability $\min(1, AR(\theta', \theta_{k-1}^{(i)}; \phi_k, \mathbf{y}^{(i)}, \mathbf{y}'))$, where

$$\begin{aligned} AR(\theta', \theta_{k-1}^{(i)}; \phi_k, \mathbf{y}^{(i)}, \mathbf{y}') &= \frac{\pi(\theta'|\phi_k; \mathbf{y}^{(i)})}{\pi(\theta_{k-1}^{(i)}|\phi_k; \mathbf{y}^{(i)})} \cdot \frac{\pi(\mathbf{y}'|\theta_{k-1}^{(i)})}{\pi(\mathbf{y}'|\theta')} \\ &= \frac{\pi(\mathbf{y}^{(i)}|\theta')}{\pi(\mathbf{y}^{(i)}|\theta_{k-1}^{(i)})} \frac{\pi(\theta'|\phi_k)}{\pi(\theta_{k-1}^{(i)}|\phi_k)} \cdot \frac{\pi(\mathbf{y}'|\theta_{k-1}^{(i)})}{\pi(\mathbf{y}'|\theta')} \\ &= \exp\left\{[\theta' - \theta_{k-1}^{(i)}]^T [s(\mathbf{y}^{(i)}) - s(\mathbf{y}')]\right\} \cdot \frac{\pi(\theta'|\phi_k)}{\pi(\theta_{k-1}^{(i)}|\phi_k)} \end{aligned} \tag{2.20}$$

set $\theta_k^{(i)} = \theta'$. Otherwise, set $\theta_k^{(i)} = \theta_{k-1}^{(i)}$.

The full exchange-within-Gibbs scheme is outlined in Algorithm 2.3. Since each step samples from the respective full conditional distribution, the algorithm ensures that the stationary distribution of the resulting Markov chain is indeed the joint posterior $\pi(\boldsymbol{\theta}, \phi|\boldsymbol{y})$ [148]. As with the exchange algorithm for the single Bayesian ERGM, the most computationally expensive step is sampling $\mathbf{y}'$ from $\pi(\cdot|\theta')$, i.e. simulating an exponential random graph with parameter θ'. Moreover, this step must be performed for each of the individual-level parameter $\theta^{(i)}$ updates. Thus, the computational cost of each iteration increases linearly with the number of networks in the data. However, since (conditional on ϕ) the $\theta^{(i)}$ are independent, these updates may be performed in parallel. Therefore, with access to a sufficient number of computing cores, the actual computational time per iteration need not necessarily increase linearly with the number of networks.

Even with access to a computing cluster, the exchange-within-Gibbs algorithm remains highly computationally intensive. Improving the efficiency of the algorithm will prove crucial in order to fit the model to data in a reasonable amount of time. We discuss the choice of proposal function h and initial values $(\phi_0, \theta_0^{(1)}, \ldots, \theta_0^{(n)})$, as well as a number of adaptations to the algorithm, in Section 2.5.

Extensions

While the above model describes a single set of networks belonging to a single population, it is easy to see how this framework could be adapted to deal with further levels of hierarchy. For example, suppose instead that we have J groups of n networks each, writing $\mathbf{Y}^{(i,j)}$ for the i^{th} network in the j^{th} group. Then, we may specify different *group-level* prior

Algorithm 2.3 The exchange-within-Gibbs algorithm for a multilevel Bayesian ERGM

Require: number of MCMC iterations K, initial values $(\phi_0, \theta_0^{(1)}, \ldots, \theta_0^{(n)})$

for $k = 1, \ldots, K$ **do**

- draw $\phi_k \sim \pi(\cdot|\boldsymbol{\theta}_{k-1})$

for $i = 1, \ldots, n$ **do**

- draw $\theta' \sim h(\cdot|\theta_{k-1}^{(i)})$
- draw $\mathbf{y}' \sim \pi(\cdot|\theta')$
- set $\theta_k^{(i)} = \theta'$ with probability

$$\min\left(1, AR(\theta', \theta_{k-1}^{(i)}; \phi_k, \mathbf{y}^{(i)}, \mathbf{y}')\right) \qquad \triangleright \text{ See Eq. (2.20)}$$

- else, set $\theta_k^{(i)} = \theta_{k-1}$

end for

end for

distributions for networks within the same group, i.e.

$$\begin{aligned} \mathbf{Y}^{(i,j)} &\sim \pi(\cdot|\theta^{(i,j)}), \quad i = 1, \ldots, n, \quad j = 1, \ldots, J \\ \theta^{(i,j)} &\sim \pi(\cdot|\phi^{(j)}), \quad i = 1, \ldots, n, \quad j = 1, \ldots, J \\ \phi^{(j)} &\sim \pi(\cdot|\beta), \quad j = 1, \ldots, J \end{aligned} \tag{2.21}$$

where β is now the population-level hyperparameter and the $\phi^{(j)}$ are group-level hyperparameters. This model is depicted in Figure 2.2 for the case of $J = 2$.

The above example is particularly relevant to neuroimaging studies in which the goal is to compare the connectivity structure of the brain between groups of individuals. By modelling the entire set of networks in this way, one can characterise the connectivity structure at a group level through the posterior distributions of the group-level parameters $\phi^{(1)}, \ldots, \phi^{(J)}$. Furthermore, we can assess group differences in the global network structure by comparing these group-level posteriors. We shall see such an application in Chapter 3, but for the remainder of this chapter, we will assume that all individuals come from a single group.

2.3.1 Synthetic dataset

In this section, we introduce a synthetic dataset of networks. The purpose of this dataset is threefold: to provide a concrete example of a population of networks; to motivate some

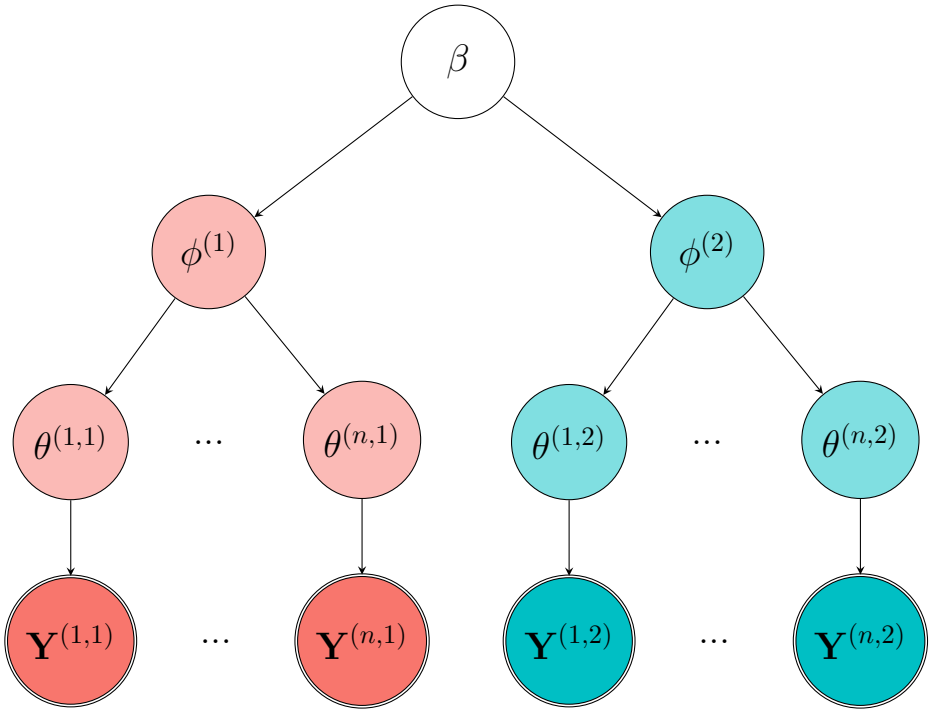


Figure 2.2: Depending on the structure of our data, we may easily add further layers to the multilevel framework. Here, we have two groups of n networks. Each network $\mathbf{Y}^{(i,j)}$ is modelled as an exponential random graph with individual-level parameter $\theta^{(i,j)}$. In turn, each individual-level parameter $\theta^{(i,j)}$ is assumed to be drawn from a group-level distribution with group-level hyperparameter $\phi^{(j)}$. Finally, the group-level hyperparameters $\phi^{(j)}$ are assumed to be drawn from a population-level distribution with hyperparameter β.

of the analysis decisions in Section 2.3; and to assess the fitting procedure with a known truth. The synthetic dataset is inspired by the brain connectivity networks that we will analyse in greater detail in Chapter 3. The procedure for constructing these networks is outlined in Section 3.2.

To generate the synthetic dataset, we simulated exponential random graphs based on these brain networks. To do so, we first fit a single Bayesian ERGM to each of the $N_{pop} = 652$ networks from the Cam-CAN dataset, using the `Bergm` package [25] in `R`, with the following three summary statistics: number of edges (edges), geometrically-weighted edgewise shared partner (GWESP), geometrically-weighted non-edgewise share partner (GWNSP). Descriptions of these statistics can be found in Table 2.1. For each network $\mathbf{y}^{(i)}$, we generated $K = 10000$ posterior samples $\{\theta_k^{(i)}\}_{k=1}^K$ from its respective posterior distribution $\pi(\theta^{(i)}|\mathbf{y}^{(i)})$ and calculated a posterior mean estimate

$$\bar{\theta}^{(i)} = \frac{1}{K}\sum_{k=1}^{K}\theta_k^{(i)}.$$

From these posterior mean estimates, we calculated a group mean $\bar{\mu}_{pop} = \frac{1}{N_{pop}}\sum_{i=1}^{N_{pop}}\bar{\theta}^{(i)}$ and group covariance $\hat{\mathbf{\Sigma}}_{pop} = \frac{1}{N_{pop}-1}\sum_{i=1}^{N_{pop}}(\bar{\theta}^{(i)} - \bar{\mu}_{pop})(\bar{\theta}^{(i)} - \bar{\mu}_{pop})^T$. As the ground truth for the synthetic dataset, we set $\mu_{true} = \bar{\mu}_{pop}$ and $\mathbf{\Sigma}_{\theta,true} = \hat{\mathbf{\Sigma}}_{pop}/10$, from which we simulate using our multilevel model. The downscaling by ten of the covariance matrix was performed to ensure a coherent single group within the synthetic dataset; a priori we believe there are distinct subgroups within the Cam-CAN data, meaning $\hat{\mathbf{\Sigma}}_{pop}$ will lead to over-dispersion relative to our intended group covariance.

To generate a population of n networks, we first generate n individual-level parameters $\theta^{(i)} \sim \mathcal{N}(\mu_{true}, \Sigma_{\theta,true})$. Then, for each $\theta^{(i)}$, we simulated a single exponential random graph from the model. This yields n networks with a similar structure to real brain connectivity networks. Since we know the parameter values that generated the networks, we may use this dataset to assess any model fitting procedure.

2.4 Model assessment

To evaluate our framework, we investigate three aspects of the model: the specification of the prior distributions (Section 2.4.2), the convergence of the fitting procedure (Section 2.4.3), and the goodness-of-fit derived from the resulting posterior samples (Sec-

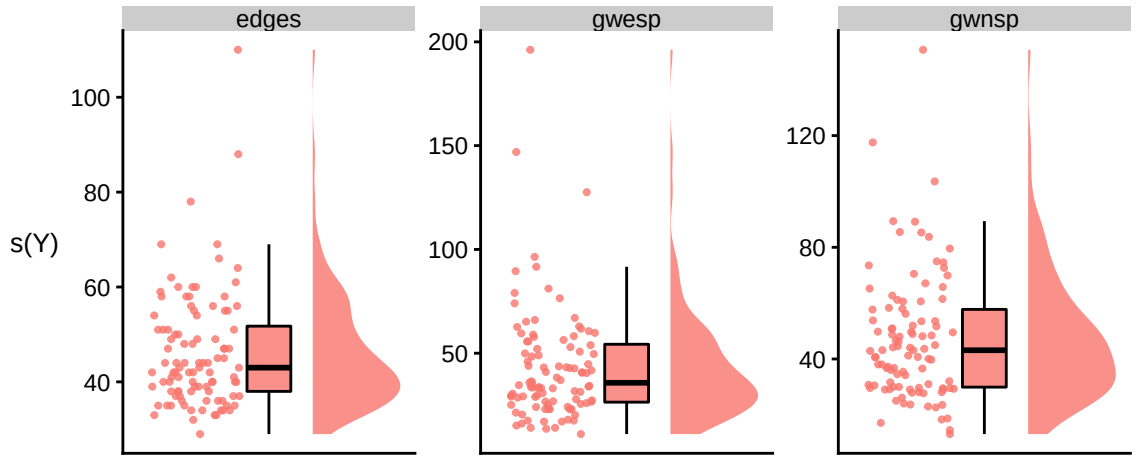


Figure 2.3: A raincloud plot is a combination of a scatter plot, box plot, and density plot. Each dot in the scatter plot corresponds to a single observation, in this case, a network. The box plots depict the median (horizontal line), the interquartile range (box), and the most extreme values up to 1.5 times the interquartile range (vertical lines). The smoothed density plots augment the box plots by providing a more detailed view of the distribution of the data.

tion 2.4.4). To assess both the prior and the posterior, we make use of predictive checks. This necessitates the comparison of populations of networks, a somewhat challenging task. To this end, we make use of raincloud plots, introduced in Section 2.4.1.

2.4.1 Raincloud plots

To visualise the distribution of graph metrics of a population of networks, we use raincloud plots [8]. A raincloud plot is a combination of a scatter plot, box plot and density plot, and can be used to visualise the distribution of any univariate metric. Each dot in the scatter plot corresponds to a single observation, in this case, a network. The box plots depict the median (horizontal line), the interquartile range (box), and the most extreme values up to 1.5 times the interquartile range (vertical lines). The smoothed density plots augment the box plots by providing a more detailed view of the distribution of the data. See Figure 2.3 for a raincloud plot of the observed summary statistics (edges, GWESP, GWNSP) of the synthetic population of networks described in Section 2.3.1.

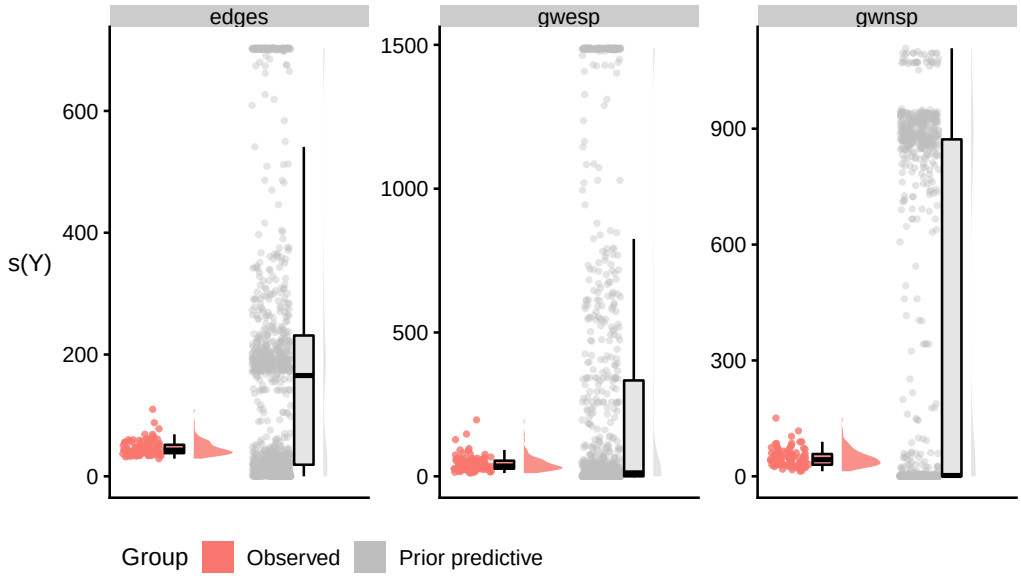


Figure 2.4: Prior predictive check for μ. The red plot corresponds to the $n = 100$ networks of the synthetic dataset (see Section 2.3.1). The grey plot corresponds to 1000 networks generated under the prior predictive distribution. The range of prior predictive values covers those of the synthetic data, indicating an appropriate prior distribution.

2.4.2 Prior predictive checks

Given the complex geometry of exponential random graph models [115], it is important to verify that the prior distributions are appropriately specified. To do so, we perform prior predictive checks by simulating (populations of) networks from the prior and comparing them graphically to the observed data. Although this may be relatively straightforward for simple examples, it is rather more difficult for our model given its hierarchical structure, and the challenge of comparing networks.

We check the prior specification in two stages. First, we assess whether the prior for the group-level mean parameter μ generates exponential random graphs that are broadly similar to the observed networks in terms of the model summary statistics. Secondly, we check the prior for the group-level covariance parameter by comparing the variance of the observed summary statistics to populations of networks generated under the prior distribution. We now describe these two steps in greater detail.

To determine the suitability of the prior for the mean parameter μ, we generate 1000

samples from the prior predictive distribution for a single network:

$$\pi_\mu(\mathbf{y}) \sim \int \pi(\mathbf{y}|\mu)\pi(\mu)d\mu \tag{2.22}$$

To do so, we first sample from the prior $\mu_i \sim \pi(\mu)$ and then simulate a single exponential random graph $\mathbf{Y}_i \sim ERGM(\mu_i)$.

To compare observed networks with the networks simulated from the prior predictive distribution, we use raincloud plots to visualise the distribution of the summary statistics of the respective networks. Figure 2.4 depicts the summary statistics of the synthetic dataset next to those of 1000 networks simulated from the prior predictive distribution under the conditionally conjugate prior $\mu \sim \mathcal{N}(0, 100I)$. The networks generated under the prior predictive distribution exhibit a wide range of summary statistic values that cover those in the synthetic dataset, indicating that the prior is appropriately specified. Observe that the complexity of ERGMs is evidenced by the prior predictive distribution: a spherical prior in parameter space appears to correspond to a multimodal distribution in network space.

To check the suitability of the prior on the covariance parameter $\mathbf{\Sigma}_\theta$, we generate sets of networks from the prior predictive distribution for a *population* of networks *conditional on the hyperparameter* μ_0:

$$\pi_\mu(\boldsymbol{y}) \sim \int \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.23}$$

This is done by first simulating $\mathbf{\Sigma}_\theta \sim \mathcal{W}^{-1}(\nu_\theta, \mathbf{\Psi}_\theta)$. We then generate n individual-level parameters $\tilde{\theta}_i \sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_\theta)$ and, from each of these, generate an exponential random graph $\mathbf{Y}_i \sim ERGM(\theta_i)$. This process results in a population of networks that we can contrast with the observed data.

Due to space limitations, it is difficult to plot more than a handful of simulated populations at once. As a result, we only plot up to 8 populations (along with the observed data) selected to reflect the possible variability of the summary statistics within the populations generated under the prior (Figure 2.5). In this example, we set $\mu_0 = \mu_{true}$. The majority of the networks in the prior predictive populations (grey plots) have similar summary statistics to those of the synthetic data (red plot). Some of the populations, however, also contain networks with vastly different summary statistics, suggesting the prior is not overinformative.

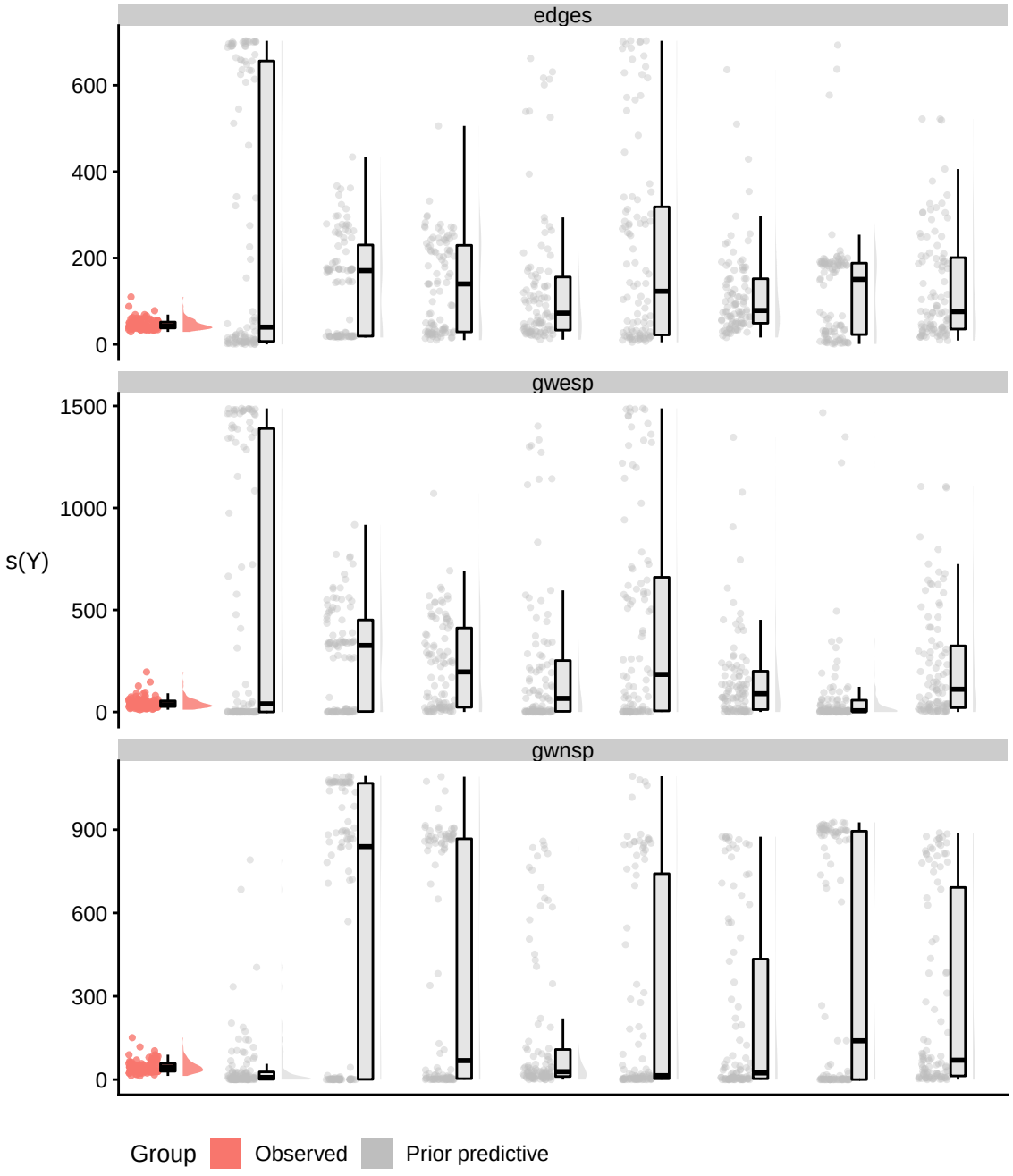


Figure 2.5: Prior predictive check for $\mathbf{\Sigma}_\theta$. The red plot corresponds to the $n = 100$ networks of the synthetic dataset (see Section 2.3.1). The eight grey plots correspond to eight populations of $n = 100$ networks generated under the prior predictive distribution conditional on a fixed mean $\mu_0 = \mu_{true}$. While the majority of the networks in the prior predictive populations have similar summary statistics to those of the synthetic data, some also contain networks with vastly different summary statistics, suggesting the prior is not overinformative.

2.4.3 Convergence

The exchange-within-Gibbs algorithm produces a sequence of samples of the model parameters $(\boldsymbol{\theta}, \phi)$ with the correct stationary distribution $\pi(\boldsymbol{\theta}, \phi|\boldsymbol{y})$. Before we can reliably use these samples, however, we must first check that the sequence has converged. A key tuning parameter in the algorithm is the number of MCMC iterations K. The aim is to choose K large enough such that we can be confident that the sequence has converged, and have enough draws from the posterior to reliably estimate any quantity of interest.

We assess convergence visually, using trace plots. By plotting the sequence of samples produced by an MCMC algorithm for a given model parameter, trace plots provide a way to check whether a given sequence has reached stationarity. We illustrate this with an example of a Bayesian ERGM fit to a single 38-node network from the synthetic dataset described in Section 2.3.1. To recap, this network was generated from an ERGM with three summary statistics (edges, geometrically weighted edgewise shared partners, and geometrically weighted non-edgewise shared partners) with corresponding parameter values inspired by real brain connectivity networks. The generating parameter values for this particular network were $\theta = (-1.94, 0.69, -0.49)$. Figure 2.6a displays the trace plot for the sequence of θ values produced by Algorithm 2.2.

In this example, we initialised the algorithm at $\theta_0 = (-18.13, -9.31, -13.60)$, which was generated at random from $\mathcal{N}(0, 100I)$. For the first 2000 iterations or so, the sequence of samples appears to explore the parameter space before settling on the target distribution $\pi(\theta|\mathbf{y})$. This initial period of iterations, known as *burn-in*, is highly influenced by the choice of starting values. To reduce their impact, we discard those iterations before the sequence appears to have converged. Figure 2.6b shows the trace plot for the same sequence of samples after discarding the first 5000 iterations.

Even if the sequence of samples produced by a model fitting procedure appears to have converged, this does not ensure that the samples are being drawn from the target distribution. For example, the sequence may be stuck in a local maximum of the distribution. In the case of simulated data, we may check whether the distribution of the samples indeed covers the (known) true model parameter values. The density plots in Figure 2.7 confirm this in the case of the single Bayesian ERGM fit.

Once we are confident that we are drawing samples from the posterior distribution, we may compute the *effective sample size* of the sequence for a given (univariate) parameter. If the samples from the algorithm were independent, then the effective sample size would simply be the number of iterations. However, in practice we have to account for the

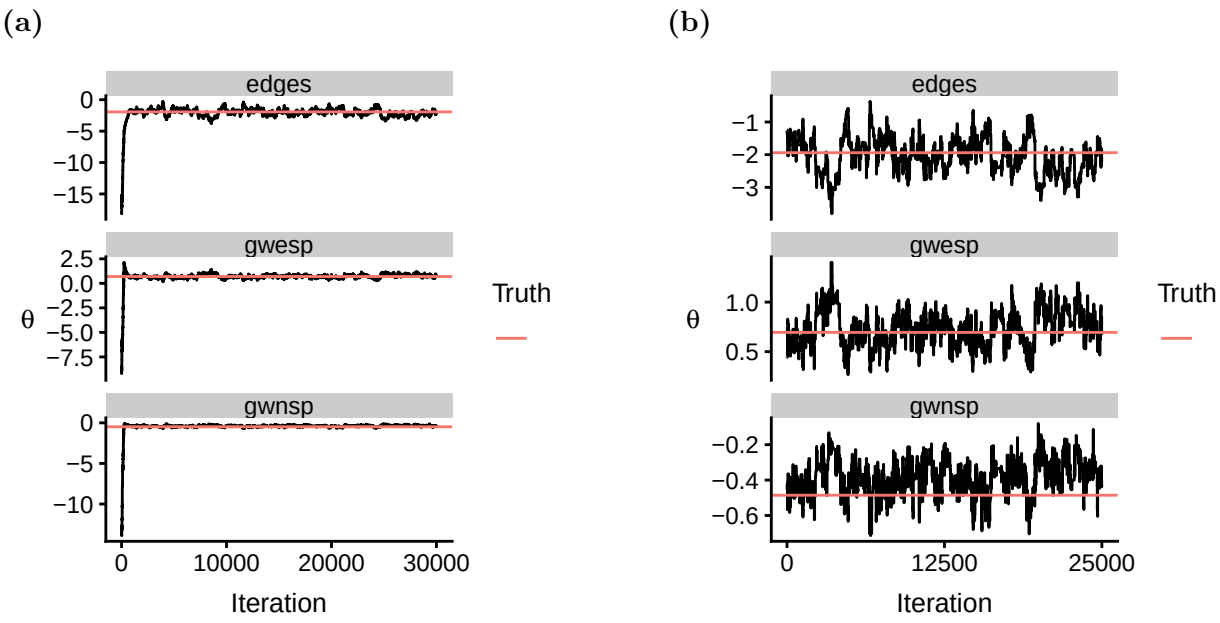


Figure 2.6: (a) Trace plot depicting the sequence of samples produced by the model fitting procedure for a single Bayesian ERGM on a simulated network. After about 2500 iterations, the sequence of samples appears to converge around the true underlying parameter values. (b) The same trace plot after discarding the first 5000 iterations.

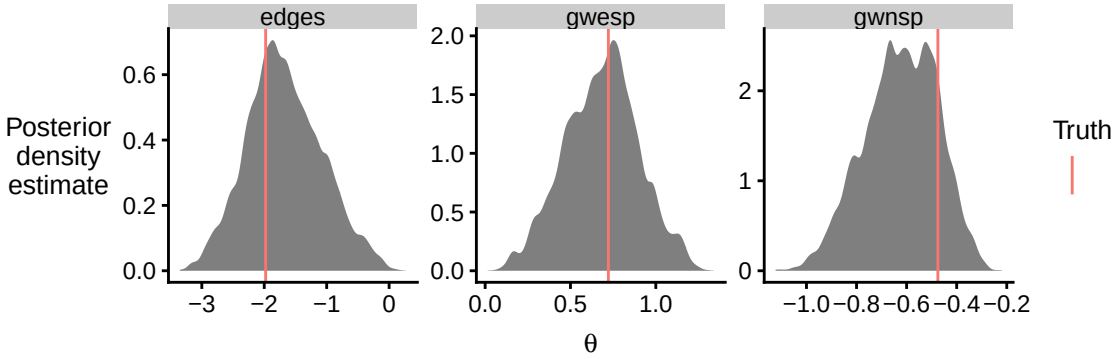


Figure 2.7: The posterior density plot illustrates the kernel density estimate for the sequence of samples produced by the fitting algorithm. In this case, the true values for all three model parameters are covered by the posterior density, suggesting that the samples are being drawn from the correct target distribution.

autocorrelation of the sequence of samples. To do so, we use the `effectiveSize` function in the R package `coda` [110]. This estimates the effective sample size N^* for a univariate parameter ψ as:

$$N^* = N\frac{\hat{\sigma}^2_\psi}{\hat{\rho}_\psi} \tag{2.24}$$

where $\hat{\sigma}^2_\psi$ is the variance of the sequence of ψ samples, $\hat{\rho}_\psi$ is an estimate of the spectral density at frequency zero of the sequence, and N is the number of MCMC iterations (after burn-in). A sequence with high autocorrelation will have a large value of $\frac{\hat{\sigma}^2_\psi}{\hat{\rho}_\psi}$ and so a comparably small effective sample size. Since each iteration of Algorithm 2.3 is computationally expensive, we would therefore prefer to construct sequences with low autocorrelation, in order to increase the effective sample size per MCMC iteration. We explore various approaches towards reducing autocorrelation in Section 2.5.

2.4.4 Posterior predictive checks

Having produced a sufficient number of samples from the posterior distribution, we then assess whether the model adequately describes the data. Since determining the distribution of appropriate test quantities is difficult, assessing such goodness-of-fit for ERGMs is typically performed graphically [76]. For a single ERGM fit, one can simulate a large number of networks from the fitted model and compare these simulations to the observed network. This comparison is usually done via a set of network metrics. If a model fits the data well then the network metrics of the simulations should be similar to those of the observed network.

Ideally, these metrics should reflect properties of the data that are deemed to be important. For example, these should include the summary statistics already in the model. We will also assess goodness-of-fit for a range of network measures commonly used in brain connectivity studies, namely (degree) assortativity, transitivity and average path length [121].

For a *population* of networks, we can apply the same principles. Instead of simulating single networks, however, we will simulate entire populations of networks from the model fit and compare network metrics of the simulated populations to the observed population. For the model fitting procedure described in Section 2.3, we consider two possible approaches to simulating a population of networks. The first approach, individual-level resampling, is based on the samples for the individual-level parameters $\theta^{(1)}, \ldots, \theta^{(n)}$. Each individual-level parameter corresponds to a single network in the observed population of

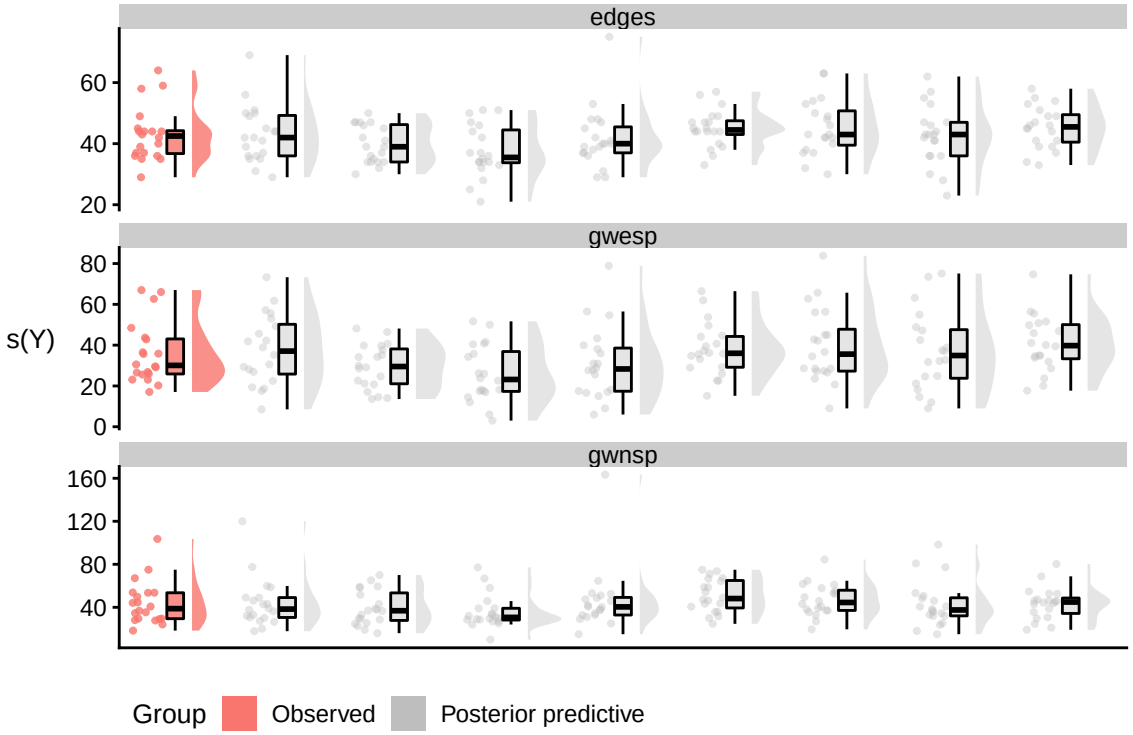


Figure 2.8: To assess model fit, we can compare network metrics of the observed networks (red) to those of simulated populations of networks (grey). The simulated networks are generated from ERGMs with parameter values drawn from the samples produced by the model fitting procedure, using an individual-level resampling approach.

networks. Thus, we may pick an iteration k uniformly at random and simulate networks $y^{(i)'} \sim \pi(\cdot|\theta_k^{(i)})$ for $i = 1, \ldots, n$.

The second approach, group-level resampling, involves re-simulating n individual-level parameters based on the samples for the group-level parameters $(\mu, \mathbf{\Sigma}_\theta)$. We again choose an iteration k uniformly at random. We then draw $\theta^{(i)'} \sim \mathcal{N}(\mu_k, \mathbf{\Sigma}_{\theta,k})$ for $i = 1, \ldots, n$ and use these values to simulate networks $y^{(i)'} \sim \pi(\cdot|\theta_k^{(i)'})$. For both approaches, we can repeat this procedure in order to construct multiple simulated populations of networks.

Visualising the distribution of network metrics for a population is somewhat more challenging than for a single network. One way is to use raincloud plots to compare the observed population metrics to a number of simulated population metrics (Figure 2.8). In particular, these plots allow us to ascertain whether the range and mean of network metric values are comparable across the observed and the simulated populations.

The raincloud plots provide a simple and rudimentary view of the distribution of the simulated population metrics, and can be useful to check if they at least loosely reflect the observed population metrics. However, they are limited to only visualising relatively few simulated populations at a time. One alternative is to graph the posterior predictive distribution of relevant test statistics [56]. For example, we could compare the mean of various network metrics across the population (Figure 2.9). To obtain an approximate posterior predictive distribution, we can simulate, say, 100 populations of networks and compute the mean of the network metrics for each of these populations. For an adequate model fit, we would expect the network metric means for the observed population to be covered by the histogram of the values derived from the simulated populations.

2.5 Computational efficiency

Sampling from the posterior of the multilevel model described in Section 2.3 is a computationally intensive procedure. This is largely due to the 'exchange' step of the algorithm, which necessitates simulating an exponential random graph (a computationally expensive operation) at each iteration. In order to improve the computational efficiency of the fitting procedure, we explore a number of adaptations to the algorithm. Section 2.5.1 introduces an alternative parametrisation of the model; Section 2.5.2 discusses the choice of proposal function in the exchange step; Section 2.5.3 addresses the initialisation of the fitting procedure; Section 2.5.4 considers the choice of prior; and Section 2.5.5 describes some further exploratory adaptations that we considered but were not included in the

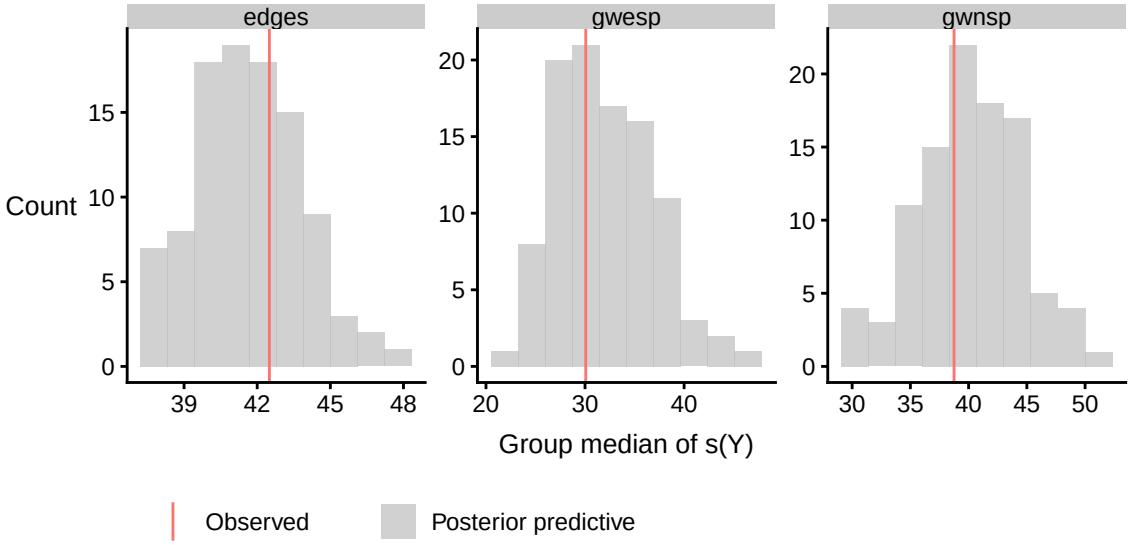


Figure 2.9: Graphical posterior predictive checks are another alternative to assessing model fit. The posterior predictive distributions of the population mean of various network metrics can be approximated by the histogram of a large number (e.g. 100) of simulated populations of networks. We can then compare the population mean of the observed network metrics (red lines) to the respective distributions.

final fitting procedure.

2.5.1 Parametrisation

The parametrisation of the multilevel model presented in Section 2.3 is known as a *centred* parametrisation. The parametrisation of general multilevel models in the context of MCMC computation has been studied in some detail [55, 106, 107, 167]. In particular, and as we shall see for our own model, the parametrisation can have a significant effect on the performance of the model-fitting procedure. Here, we discuss and compare the centred and the non-centred parametrisations. We also describe the 'interweaving' strategy [167], which combines both parametrisations.

Recall the model presented in Section 2.3. At level of the data, each network is modelled as an exponential random graph, $\mathbf{Y}^{(i)} \sim \pi(\cdot|\theta^{(i)})$. Next, each individual-level parameter is assumed to be drawn from a common group-level distribution $\theta^{(i)} \sim \mathcal{N}(\mu, \mathbf{\Sigma}_\theta)$. This is known as a centred parametrisation (CP) [55, 107], in which the group-level parameters $(\mu, \mathbf{\Sigma}_\theta)$ are independent of the data $\boldsymbol{Y}$. Ignoring the priors for the time being,

we can write this as:

$$\begin{aligned} \mathbf{Y}^{(i)} &\sim ERGM(\theta^{(i)}), \quad i = 1, \ldots, n \\ \theta^{(i)} &\sim \mathcal{N}(\mu, \mathbf{\Sigma}_\theta), \quad i = 1, \ldots, n. \end{aligned} \tag{2.25}$$

In contrast, the *non-centred* parametrisation (NCP) can be written as follows:

$$\begin{aligned} \mathbf{Y}^{(i)} &\sim ERGM(\mu + \tilde{\theta}^{(i)}), \quad i = 1, \ldots, n \\ \tilde{\theta}^{(i)} &\sim \mathcal{N}(0, \mathbf{\Sigma}_\theta), \quad i = 1, \ldots, n. \end{aligned} \tag{2.26}$$

Note that the group-level parameter μ has entered the likelihood directly in (2.26). Via the transformation $\theta^{(i)} = \mu + \tilde{\theta}^{(i)}$, we see that these two parametrisations indeed describe the same model. To implement the non-centred parametrisation, however, we will have to modify Algorithm 2.3. This is because the conditional distribution of μ given $(\boldsymbol{Y}, \boldsymbol{\theta}, \mathbf{\Sigma}_\theta)$ is no longer independent of $\boldsymbol{Y}$ and is therefore intractable:

$$\begin{aligned} \pi(\mu|\tilde{\boldsymbol{\theta}}, \mathbf{\Sigma}_\theta; \underline{\mathbf{y}}) &= \pi(\mu|\tilde{\boldsymbol{\theta}}; \boldsymbol{y}) \\ &= \frac{\pi(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}, \mu)\pi(\mu|\tilde{\boldsymbol{\theta}})}{\pi(\boldsymbol{y})} \\ &= \frac{\prod_{i=1}^n \pi(\mathbf{y}^{(i)}|\tilde{\theta}^{(i)}, \mu)\pi(\mu)}{\pi(\boldsymbol{y})} \\ &= \frac{\exp\{\sum_{i=1}^n [(\mu + \tilde{\theta}^{(i)})^T s(\mathbf{y}^{(i)})] - \frac{1}{2}(\mu - \alpha)^T \mathbf{\Sigma}_\mu^{-1}(\mu - \alpha)\}}{\pi(\boldsymbol{y})|2\pi\mathbf{\Sigma}_\mu|^{1/2} \prod_{i=1}^n Z(\mu + \tilde{\theta}^{(i)})} \end{aligned} \tag{2.27}$$

As a result, we cannot sample directly from $\pi(\mu|\tilde{\boldsymbol{\theta}}, \mathbf{\Sigma}_\theta; \boldsymbol{y})$. We need to perform an additional exchange step to update μ, as follows. At the k^{th} iteration, first draw a proposal μ' from a symmetric proposal function $h_\mu(\cdot|\mu_{k-1})$. Then, draw a population of exponential random graphs $\mathbf{y}^{(i)'}$ from $ERGM(\mu' + \theta_{k-1}^{(i)})$ for $i = 1, \ldots, n$ respectively. With probability $\min\Big(1, AR(\mu', \mu_{k-1}; \tilde{\boldsymbol{\theta}}_{k-1}, \mathbf{\Sigma}_{\theta,k}, \boldsymbol{y}, \boldsymbol{y}')\Big)$, set $\mu_k = \mu'$ where

$$
\begin{aligned}
AR(\mu', \mu_{k-1}; \tilde{\boldsymbol{\theta}}_{k-1}, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{y}, \boldsymbol{y}') &= \frac{\pi(\mu'|\tilde{\boldsymbol{\theta}}_{k-1}, \boldsymbol{\Sigma}_{\theta,k}; \boldsymbol{y})}{\pi(\mu_{k-1}|\tilde{\boldsymbol{\theta}}_{k-1}, \boldsymbol{\Sigma}_{\theta,k}; \boldsymbol{y})} \cdot \frac{\pi(\boldsymbol{y}'|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1})}{\pi(\boldsymbol{y}|\mu', \tilde{\boldsymbol{\theta}}_{k-1})} \\
&= \frac{\pi(\boldsymbol{y}|\mu', \tilde{\boldsymbol{\theta}}_{k-1})}{\pi(\boldsymbol{y}|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1})} \cdot \frac{\pi(\mu'|\boldsymbol{\Sigma}_{k-1})}{\pi(\mu_{k-1}|\boldsymbol{\Sigma}_{k-1})} \cdot \frac{\pi(\boldsymbol{y}'|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1})}{\pi(\boldsymbol{y}'|\mu', \tilde{\boldsymbol{\theta}}_{k-1})} \\
&= \exp\left\{(\mu' - \mu_{k-1}) \sum_{i=1}^{n} (\mathbf{y}^{(i)} - \mathbf{y}^{(i)'})\right\} \cdot \frac{\pi(\mu'|\boldsymbol{\Sigma}_{k-1})}{\pi(\mu_{k-1}|\boldsymbol{\Sigma}_{k-1})}
\end{aligned}
\tag{2.28}
$$

Otherwise, set $\mu_k = \mu_{k-1}$. The full MCMC procedure for the non-centred parametrisation of the multilevel model is described in Algorithm 2.4.

Algorithm 2.4 The exchange-within-Gibbs algorithm for a multilevel Bayesian ERGM: *Non-centred parametrisation*

Require: number of MCMC iterations N, initial values $(\mu_0, \tilde{\theta}_0^{(1)}, \ldots, \tilde{\theta}_0^{(n)})$

for $k = 1, \ldots, N$ **do**

- draw $\boldsymbol{\Sigma}_{\theta,k} \sim \pi(\cdot|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1})$
- draw $\mu' \sim h_\mu(\cdot|\mu_{k-1})$
- draw $\mathbf{y}^{(i)'} \sim \pi(\cdot|\mu' + \tilde{\theta}_{k-1}^{(i)})$ for $i = 1, \ldots, n$
- set $\mu_k = \mu'$ with probability
  $\min\left(1, AR(\mu', \mu_{k-1}; \tilde{\boldsymbol{\theta}}_{k-1}, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{y}, \boldsymbol{y}')\right)$ ▷ See Eq. (2.28)
- else, set $\mu_k = \mu_{k-1}$

for $i = 1, \ldots, n$ **do**

- draw $\tilde{\theta}' \sim h_\theta(\cdot|\tilde{\theta}_{k-1}^{(i)})$
- draw $\mathbf{y}' \sim \pi(\cdot|\mu_k + \tilde{\theta}')$
- set $\tilde{\theta}_k^{(i)} = \tilde{\theta}'$ with probability
  $\min\left(1, AR(\tilde{\theta}', \tilde{\theta}_{k-1}^{(i)}; \boldsymbol{\Sigma}_{\theta,k}, \mathbf{y}^{(i)}, \mathbf{y}')\right)$ ▷ See Eq. (2.20)
- else, set $\tilde{\theta}_k^{(i)} = \tilde{\theta}_{k-1}$

end for

end for

To illustrate the characteristics of each parametrisation, we analyse their behaviour on the a subset of $n = 20$ networks from the synthetic dataset described in Section

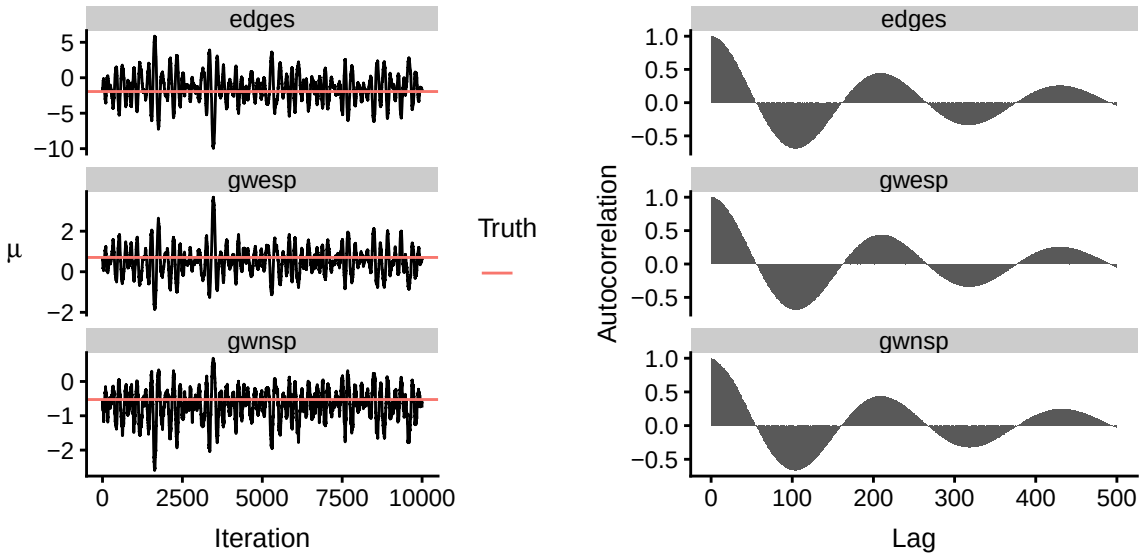


Figure 2.10: The centred parametrisation approach produces posterior samples with a strong autocorrelated pattern.

2.3.1. In each case we drew individual-level parameters $\theta^{(i)}$ from $\mathcal{N}(\mu_{true}, \Sigma_{\theta,true})$ and generated a corresponding network for each individual. To simplify the model fitting procedure, we assume here that Σ_θ is fixed and known, and focus on the performance of the parametrisations with regards to the posterior distribution of μ. To avoid any effect of burn-in, we also initialise each algorithm at the true values for μ and $\boldsymbol{\theta}$. We use a vague prior for μ of $\mathcal{N}(0, 100I)$. For the individual-level mean updates in both the centred and non-centred parametrisations, we draw proposals $\theta' \sim \mathcal{N}(0, 0.005I)$. Further, for the group-level mean updates in the non-centred parametrisation (NCP), we draw proposals $\mu' \sim \mathcal{N}(0, 0.0005I)$. We focus in particular on the sequence of μ samples.

While the sequence of samples produced by the Algorithm 2.3 (CP) appears to be centred around the underlying true values, it exhibits a strong autocorrelated pattern (Figure 2.10). This is due to the strong dependence between the group-level mean parameter μ and the individual-level parameters $\theta^{(i)}, i = 1, \ldots, n$. Under the centred parametrisation, μ is updated by drawing from

$$\mu \sim \mathcal{N}\left(\left(\boldsymbol{\Sigma}_\mu^{-1} + n\boldsymbol{\Sigma}_\theta^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_\mu^{-1}\alpha + n\boldsymbol{\Sigma}_\theta^{(-1)}\bar{\theta}\right), \left(\boldsymbol{\Sigma}_\mu^{-1} + n\boldsymbol{\Sigma}_\theta^{-1}\right)^{-1}\right)$$

where $\mathcal{N}(\alpha, \Sigma_\mu)$ is the prior for μ. In the above setting, $\alpha = 0$ and Σ_μ^{-1} is small relative to Σ_θ^{-1}, and so the update can be seen as an approximate draw from $\mathcal{N}(\bar{\theta}, \boldsymbol{\Sigma}_\theta)$.

The dependence of the $\theta^{(i)}$ updates on μ is slightly more complex. For a proposal θ', we generate an exponential random graph model $\mathbf{y}'$, and accept the proposal with probability

$$\min\left(1, \frac{\exp\left\{\theta'^T[s(\mathbf{y}^{(i)}) - s(\mathbf{y}')] - \frac{1}{2}(\theta' - \mu)^T\boldsymbol{\Sigma}_\theta^{-1}(\theta' - \mu)\right\}}{\exp\left\{\theta^{(i)T}[s(\mathbf{y}^{(i)}) - s(\mathbf{y}')] - \frac{1}{2}(\theta^{(i)} - \mu)^T\boldsymbol{\Sigma}_\theta^{-1}(\theta^{(i)} - \mu)\right\}}\right). \tag{2.29}$$

Observe that, if $s(\mathbf{y}^{(i)}) - s(\mathbf{y}')$ is small, then the acceptance probability will largely be determined by the relative proximity of θ' and $\theta^{(i)}$ to μ. If θ' is closer than $\theta^{(i)}$ to μ, then we are more likely to accept it. This leads to the co-dependence between μ and the $\theta^{(i)}, i = 1, \ldots, n$. If the μ parameter moves in one direction, the $\theta^{(i)}$ are more likely to move in this same direction.

In contrast, the NCP algorithm produces samples with rather different behaviour (Figure 2.11). While there is still clear autocorrelation in the sequence, the structure of this is much less regular compared to that of the CP approach. Further, although the variance of these samples is much smaller, they explore the parameter space much more slowly. We now provide a heuristic explanation for this behaviour. For a single proposal μ', we must generate an entire population of n networks $\mathbf{y}^{(i)'}$ from $\mu' + \tilde{\theta}^{(i)}$ respectively. We then accept the proposal with probability

$$\min\left(1, \frac{\exp\left\{\mu'^T\sum_{i=1}^n[s(\mathbf{y}^{(i)}) - s(\mathbf{y}^{(i)'})] - \frac{1}{2}\mu'^T\boldsymbol{\Sigma}_\mu^{-1}\mu'\right\}}{\exp\left\{\mu^T\sum_{i=1}^n[s(\mathbf{y}^{(i)}) - s(\mathbf{y}^{(i)'})] - \frac{1}{2}\mu^T\boldsymbol{\Sigma}_\mu^{-1}\mu\right\}}\right) \tag{2.30}$$

Suppose the jump from μ to μ' is such that exactly one of the $\mu' + \tilde{\theta}^{(i)}$ is outside of the high density region of the posterior for a single network. In turn, this will yield a simulated network that is very different to the observed network. In other words, $|s(\mathbf{y}^{(i)}) - s(\mathbf{y}^{(i)'})|$ will be very large relative to $|s(\mathbf{y}^{(j)}) - s(\mathbf{y}^{(j)'})|$ for $j \neq i$. In particular, the term $s(\mathbf{y}^{(i)}) - s(\mathbf{y}^{(i)'})$ will dominate the expression in (2.30), leading to a very small acceptance probability. Thus, large jumps are likely to be rejected.

In general, the centred parametrisation and the non-centred parametrisation tend to be complementary: when one performs poorly, the other tends to perform much better [107]. In our case, neither parametrisation seems to work particularly well. However, it is possible to combine both approaches using an ancillarity-sufficiency interweaving strategy (ASIS) [167].

ASIS works by combining the updating schemes of the CP and NCP approaches. We

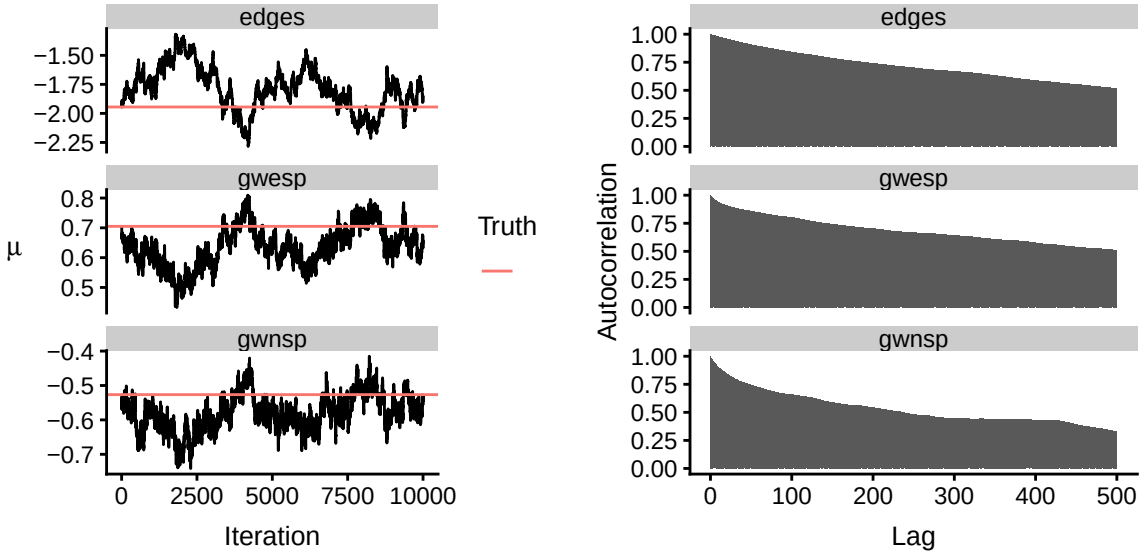


Figure 2.11: The non-centred parametrisation approach produces posterior samples with a much smaller variance than the centred-parametrisation approach. However, the chain does not appear to be mixing well.

now describe it in the context of the multilevel model fitting procedure. The updating step for Σ_θ remains the same, but we introduce an intermediate step to update $(\mu, \boldsymbol{\theta})$. Specifically, at the k^{th} iteration, we first draw intermediate updates of the individual-level parameters by performing an exchange step. For each i, draw a proposal θ' from $h_\theta(\cdot|\theta^{(i)}_{k-1})$ followed by an exponential random graph $\mathbf{y}'$ from $\pi(\cdot|\theta')$ and then set $\theta^{(i)}_{k-0.5} = \theta'$ with probability

$$\min\left(1, \frac{\exp\left\{\theta'^T[s(\mathbf{y}^{(i)}) - s(\mathbf{y}')] - \frac{1}{2}(\theta' - \mu_k)^T\boldsymbol{\Sigma}^{-1}_{\theta,k}(\theta' - \mu_k)\right\}}{\exp\left\{\theta^{(i)T}_{k-1}[s(\mathbf{y}^{(i)}) - s(\mathbf{y}')] - \frac{1}{2}(\theta^{(i)}_{k-1} - \mu_k)^T\boldsymbol{\Sigma}^{-1}_{\theta,k}(\theta^{(i)}_{k-1} - \mu_k)\right\}}\right) \tag{2.31}$$

Else, set $\theta^{(i)}_{k-0.5} = \theta^{(i)}_{k-1}$. Next, we perform an intermediate CP update of the group-level parameter μ based on these values, i.e we draw

$$\mu_{k-0.5} \sim \mathcal{N}\left(\left(\boldsymbol{\Sigma}^{-1}_\mu + n\boldsymbol{\Sigma}^{-1}_{\theta,k}\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}_\mu\alpha + n\boldsymbol{\Sigma}^{(-1)}_{\theta,k}\bar{\theta}_{k-0.5}\right), \left(\boldsymbol{\Sigma}^{-1}_\mu + n\boldsymbol{\Sigma}^{-1}_{\theta,k}\right)^{-1}\right). \tag{2.32}$$

Following this, we set $\tilde{\theta}^{(i)}_k = \theta^{(i)}_{k-0.5} - \mu_{k-0.5}$ for $i = 1, \ldots, n$. Finally, we perform an NCP update of μ. Namely, we draw a proposal μ' from $h_\mu(\cdot|\mu_{k-0.5})$ followed by a population

of exponential random graphs $\mathbf{y}^{(i')}$ from $\pi(\cdot|\mu' + \tilde{\theta}_k^{(i)})$ for $i = 1, \ldots, n$. With probability

$$\min\left(1, \frac{\exp\left\{\mu'^T \sum_{i=1}^n [s(\mathbf{y}^{(i)}) - s(\mathbf{y}^{(i)'})] - \frac{1}{2}\mu'^T \boldsymbol{\Sigma}_\mu^{-1} \mu'\right\}}{\exp\left\{\mu_{k-1}^T \sum_{i=1}^n [s(\mathbf{y}^{(i)}) - s(\mathbf{y}^{(i)'})] - \frac{1}{2}\mu_{k-1}^T \boldsymbol{\Sigma}_\mu^{-1} \mu_{k-1}\right\}}\right) \tag{2.33}$$

set $\mu_k = \mu'$. Else, set $\mu_k = \mu_{k-0.5}$. The full ASIS algorithm is described in Algorithm 2.5.

Algorithm 2.5 The ancillarity-sufficiency interweaving strategy (ASIS) algorithm for a multilevel Bayesian ERGM

Require: number of MCMC iterations N, initial values $(\mu_0, \tilde{\theta}_0^{(1)}, \ldots, \tilde{\theta}_0^{(n)})$

for $k = 1, \ldots, N$ **do**

- draw $\boldsymbol{\Sigma}_{\theta,k} \sim \pi(\cdot|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1})$

for $i = 1, \ldots, n$ **do**

- set $\theta_{k-1}^{(i)} = \mu_{k-1} + \tilde{\theta}_{k-1}^{(i)}$
- draw $\theta' \sim h_\theta(\cdot|\theta_{k-1}^{(i)})$
- draw $\mathbf{y}' \sim \pi(\cdot|\theta')$
- set $\theta_{k-0.5}^{(i)} = \theta'$ with probability
  $\min\left(1, AR(\theta', \theta_{k-1}^{(i)}; \mu_{k-1}, \boldsymbol{\Sigma}_{\theta,k}, \mathbf{y}^{(i)}, \mathbf{y}')\right)$ ▷ See Eq. (2.31)
- else, set $\theta_{k-0.5}^{(i)} = \theta_{k-1}^{(i)}$

end for

- draw $\mu_{k-0.5} \sim \pi(\cdot|\boldsymbol{\theta}_{k-0.5}, \boldsymbol{\Sigma}_{\theta,k})$

for $i = 1, \ldots, n$ **do**

- set $\tilde{\theta}_k^{(i)} = \theta_{k-0.5}^{(i)} - \mu_{k-0.5}$

end for

- draw $\mu' \sim h_\mu(\cdot|\mu_{k-0.5})$
- draw $\mathbf{y}^{(i)'} \sim \pi(\cdot|\mu' + \tilde{\theta}_k^{(i)})$ for $i = 1, \ldots, n$
- set $\mu_k = \mu'$ with probability
  $\min\left(1, AR(\mu', \mu_{k-0.5}; \tilde{\boldsymbol{\theta}}_k, \boldsymbol{\Sigma}_{\theta,k}, \mathbf{y}, \mathbf{y}')\right)$ ▷ See Eq. (2.33)
- else, set $\mu_k = \mu_{k-0.5}$

end for

We now analyse the behaviour of the ASIS algorithm on the same synthetic dataset with $n = 20$ networks. Again, we assume that Σ_θ is fixed and known, and initialise the

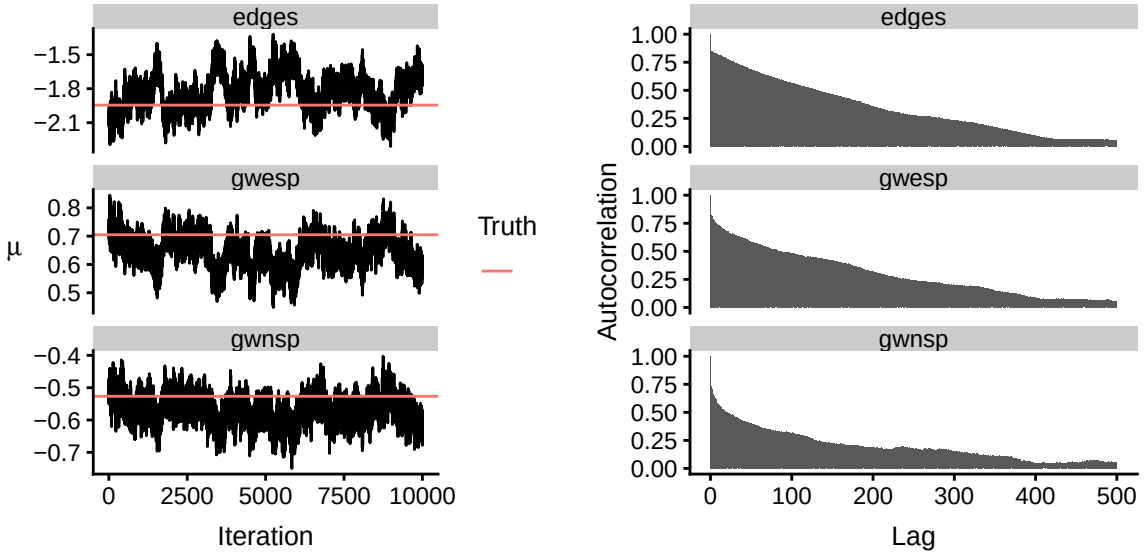


Figure 2.12: The ASIS algorithm produces posterior samples that appear to be mixing better than both the CP and the NCP approaches.

algorithm at the true values for μ and $\boldsymbol{\theta}$. We use the same vague prior for μ of $\mathcal{N}(0, 100I)$, draw proposals $\tilde{\theta}' \sim \mathcal{N}(0, 0.003I)$ and $\mu' \sim \mathcal{N}(0, 0.0005I)$. Under this setting, the ASIS algorithm appears to outperform both the CP and the NCP algorithms (Figure 2.12). The sequence of μ samples does not exhibit the same regular autocorrelation pattern produced by the CP algorithm (Figure 2.10). Moreover, it explores the parameter space more quickly than the NCP algorithm. Due to its superior performance, we will henceforth only use the ASIS algorithm to fit the multilevel model to data.

2.5.2 Proposals

The choice of proposal function in any MCMC algorithm is crucial to the efficiency of the fitting procedure. For simplicity, we will only consider multivariate normal random walk proposals. In other words, we draw proposals $\theta_{prop} \sim \mathcal{N}(\theta_{curr}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is a covariance matrix to be chosen. To choose an effective proposal function, we need to consider both the scaling and the direction of the jumps. If the jumps are too large, then the vast majority of proposals are likely to be rejected. In contrast, if the jumps are too small, then the proposals are likely to be accepted but then the sequence of samples will only explore the space very slowly.

The optimal scaling of proposal jumps for Metropolis algorithms has been investigated

in a variety of contexts [116]. In particular, it was shown that the optimal acceptance rate for a multivariate normal target distribution with independent components is approximately 0.234 in the infinite-dimensional limit, and 0.44 for a univariate normal. For a multivariate normal target distribution with correlated components, the optimal jumping rule is to use proposals with the same correlation structure as the target, with an optimal acceptance rate again of 0.234.

While they do not address proposals for the exchange algorithm, we will nevertheless use their recommendations to tune the proposals in Algorithm 2.5. In particular, we will first fix the correlation structure of the proposal jumps, and then tune the scaling of the jumps to achieve an acceptance rate of between 0.1 and 0.4. (Although 0.234 is optimal, an acceptance rate in this range will still yield an algorithm that is at least 80% efficient in the case of a multivariate normal with independent components [116].)

Ideally, the correlation structure of the proposal jumps should match the (generally, unknown) correlation structure of the posterior distribution of interest. This is particularly important for Bayesian ERGMs: the posterior distribution is often concentrated on a thin and highly correlated region [24]. One common approach is to first perform a preliminary Metropolis algorithm with uncorrelated jumps and then estimate the correlation structure based on the empirical covariance of the resulting output (see e.g. [148]). One can then run the Metropolis algorithm using jumps with the estimated correlation structure.

Since the exchange algorithm is computationally expensive, we would like to estimate the correlation structure without performing a preliminary run of the fitting procedure. We do this by first fitting a single Bayesian ERGM to each network, using the `Bergm` package [25], with the same summary statistics as in the full multilevel model. (Note that this can be done completely in parallel.) For each network $\mathbf{y}^{(i)}$, we produce $K = 12000$ samples $\{\theta_k^{(i)}\}_{k=1}^K$ from their respective posterior distributions $\pi(\theta^{(i)}|\mathbf{y}^{(i)})$. In particular, we obtain estimates $\bar{\theta}^{(i)} = \frac{1}{K}\sum_{k=1}^K \theta_k^{(i)}$ of the mean and $\hat{\mathbf{\Sigma}}^{(i)} = \frac{1}{K-1}\sum_{k=1}^K(\theta_k^{(i)} - \bar{\theta}^{(i)})(\theta_k^{(i)} - \bar{\theta}^{(i)})^T$ of the covariance for the posterior distribution from each network. We also obtain the group-level mean $\bar{\mu} = \frac{1}{n}\sum_{i=1}^n \bar{\theta}^{(i)}$.

From these individual-level estimates, we construct proposals that better explore the posterior distribution under the hierarchical framework. For the individual-level proposals, we draw $\theta' \sim \mathcal{N}(0, \tau_\theta \frac{1}{n}\sum_{i=1}^n \hat{\mathbf{\Sigma}}^{(i)})$ where $\tau_\theta > 0$ is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4, as discussed above. For the group-level proposals, we estimate the correlation structure of $\pi(\mu|\boldsymbol{\theta}, \underline{\mathbf{y}})$ by the empirical variance of these pos-

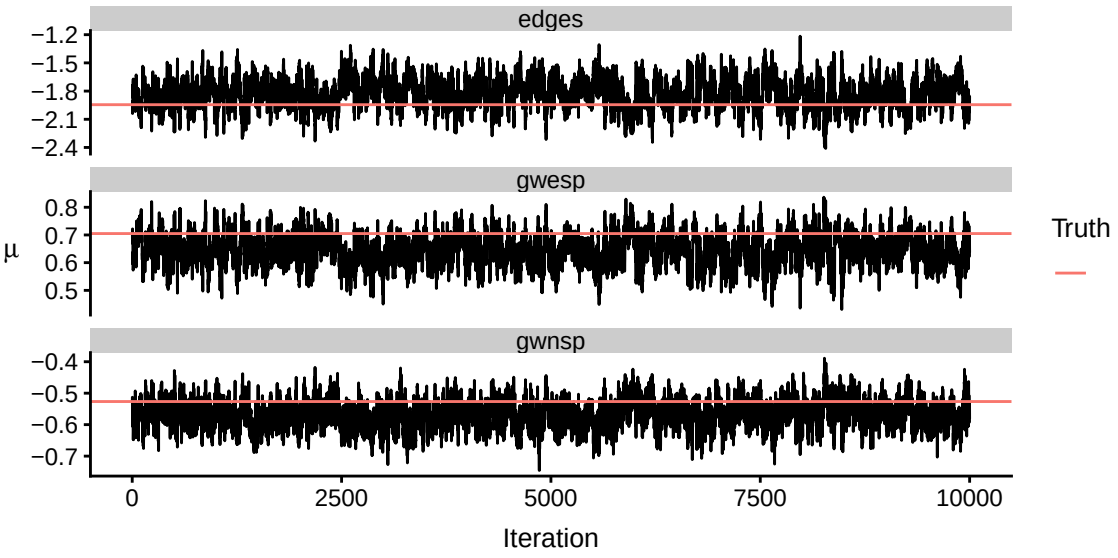


Figure 2.13: Using proposals based on the correlation structure of Bayesian ERGMs fitted independently to the individual networks leads to better mixing compared to the isotropic proposals that were used previously (Figure 2.12).

terior means. Specifically, we calculate $\mathbf{\Sigma}_{\mu,prop} = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{\theta}^{(i)} - \bar{\mu})(\bar{\theta}^{(i)} - \bar{\mu})^T$. We then draw proposals $\mu' \sim \mathcal{N}(\mu, \tau_\mu \Sigma_{\mu,prop})$ where again $\tau_\mu > 0$ is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4. For the synthetic dataset and the proposals described here, we found that setting the tuning parameters at $\tau_\theta = \tau_\mu$ led to adequate acceptance rates. Using these proposals leads to better mixing of the posterior samples (Figure 2.13). Note that our approach thus far is not empirical Bayes as we are only using the data to set the proposals and the initial values, rather than the priors (see next Section).

An alternative approach to specifying the proposal directions of a Metropolis algorithm is adaptive direction sampling [59]. This strategy, which is implemented in the `Bergm` package [25], constructs multiple interacting chains all with the same target distribution. We avoid this approach in our model fitting procedure due to the increased computational cost of constructing multiple coupled chains.

2.5.3 Initialisation

In the previous section, we initialised the algorithms at the true values for each of the parameters. Of course, in practice this is not a feasible strategy. We would, nonetheless,

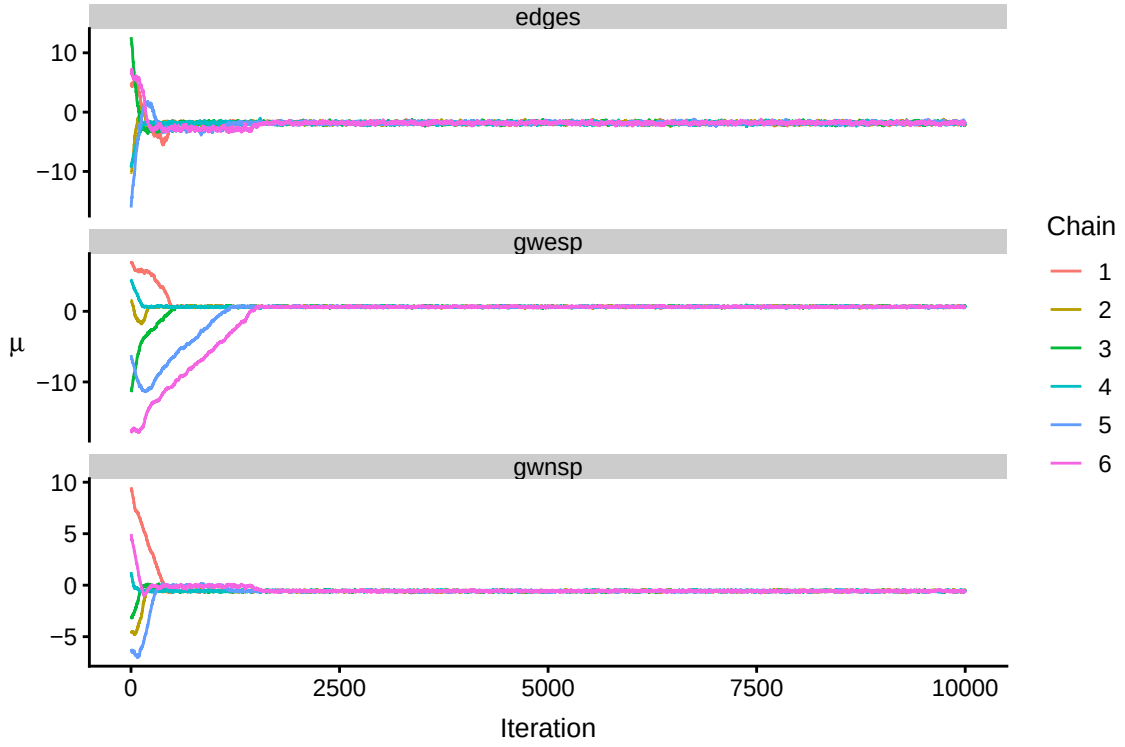


Figure 2.14: The burn-in period for the sequence of μ samples produced using Algorithm 2.5 with Σ_θ fixed and known for six chains with different initialisations. The sequences can take close to 2000 iterations to explore the parameter space before converging to the (same) target distribution.

prefer to reduce the length of the burn-in period as much as possible. If we choose poor initial starting values, the sequence of samples may take a long time to reach stationarity (Figure 2.14). We initialise the parameters in a sensible region of the parameter space by utilising the posterior samples from individual Bayesian ERGM fits to each network in the population, as outlined in the previous section. These serve as the initial values for the individual level-parameters, i.e. $(\theta_0^{(1)}, \ldots, \theta_0^{(n)}) = (\bar{\theta}^{(1)}, \ldots, \bar{\theta}^{(n)})$. Similarly, we initialise the group-level parameter μ as the mean of the individual-level posterior means, i.e. $\mu_0 = \frac{1}{n}\sum_{i=1}^{n} \bar{\theta}^{(i)}$. By using these initial values, we can effectively eliminate most of the burn-in period.

2.5.4 Priors

The choice of prior distributions for the hyperparameters μ and $\mathbf{\Sigma}_\theta$ can have a significant impact on the model fit. As described in Section 2.3, we assume a multivariate

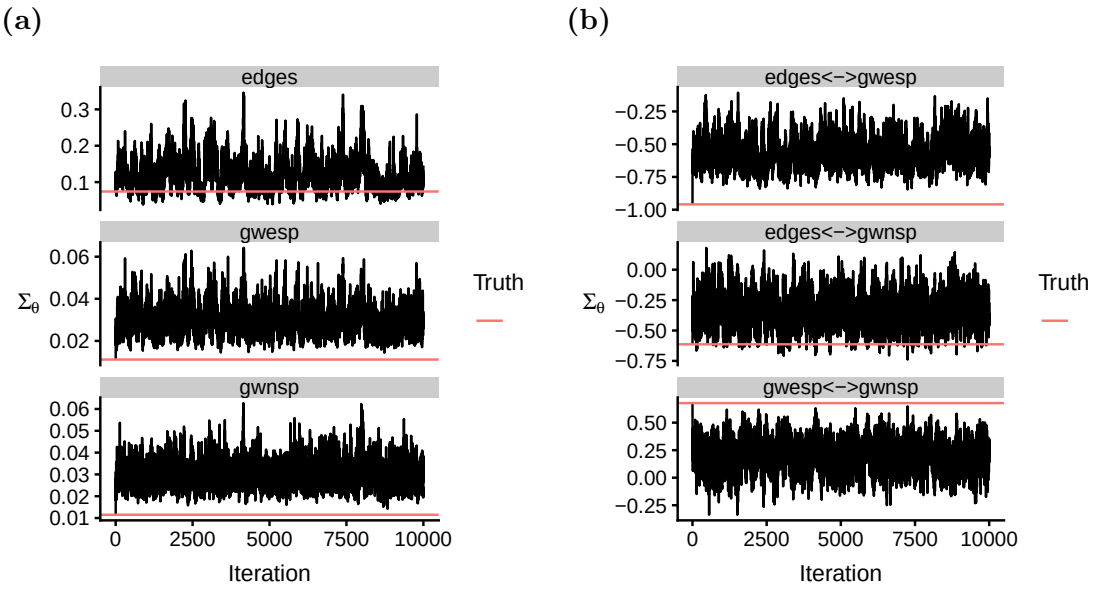


Figure 2.15: Under the vague prior, the sequence of $\mathbf{\Sigma}_\theta$ samples appear to mix well but do not capture the underlying truth for (a) the diagonal elements of $\mathbf{\Sigma}_\theta$ or (b) the correlation of the $\theta^{(i)}$ parameters, i.e. the normalised off-diagonal elements of $\mathbf{\Sigma}_\theta$.

normal prior $\mu \sim \mathcal{N}(\alpha, \mathbf{\Sigma}_\mu)$, and for $\mathbf{\Sigma}_\theta$, we assume an Inverse-Wishart distribution $\mathbf{\Sigma}_\theta \sim \mathcal{W}^{-1}(\mathbf{\Psi}, \nu)$ where $\mathbf{\Psi}$ is a constant, $p \times p$ positive-definite matrix, and $\nu \in \mathbb{R}$ is such that $\nu > p - 1$. Here, we investigate the performance of the fitting procedure and two different choices for $(\mathbf{\Psi}, \nu)$, corresponding to a 'vague' choice and an 'empirical' choice.

We first describe the 'vague' choice of prior. For the group-level mean parameter μ, this is the prior we have used thus far, namely $\mu \sim \mathcal{N}(0, 100I)$. For the group-level covariance $\mathbf{\Sigma}_\theta$, we now relax the assumption that it is fixed and known, and specify a vague prior $\mathbf{\Sigma}_\theta \sim \mathcal{W}^{-1}(I, p + 1)$ where p is the number of summary statistics in the model. We assess the performance of the fitting procedure (Algorithm 2.5) on all 100 networks from the synthetic dataset described in Section 2.3.1, with $p = 3$ summary statistics, as before.

While the algorithm produces samples for $\mathbf{\Sigma}_\theta$ that appear to mix reasonably well, they do not quite cover the underlying truth (Figure 2.15). This is reflected by the posterior predictive checks (Figure 2.16). When generated via individual-level resampling (that is, directly from the samples for the individual-level $\theta^{(i)}$ parameters), the simulated populations of networks have similar values of the summary statistics to the observed population of networks (2.16a). However, under group-level resampling (the individual-level $\theta^{(i)}$ parameters are first drawn from the samples of $(\mu, \mathbf{\Sigma}_\theta)$), the simulated populations of networks have a greater range of summary statistic values than the observed population

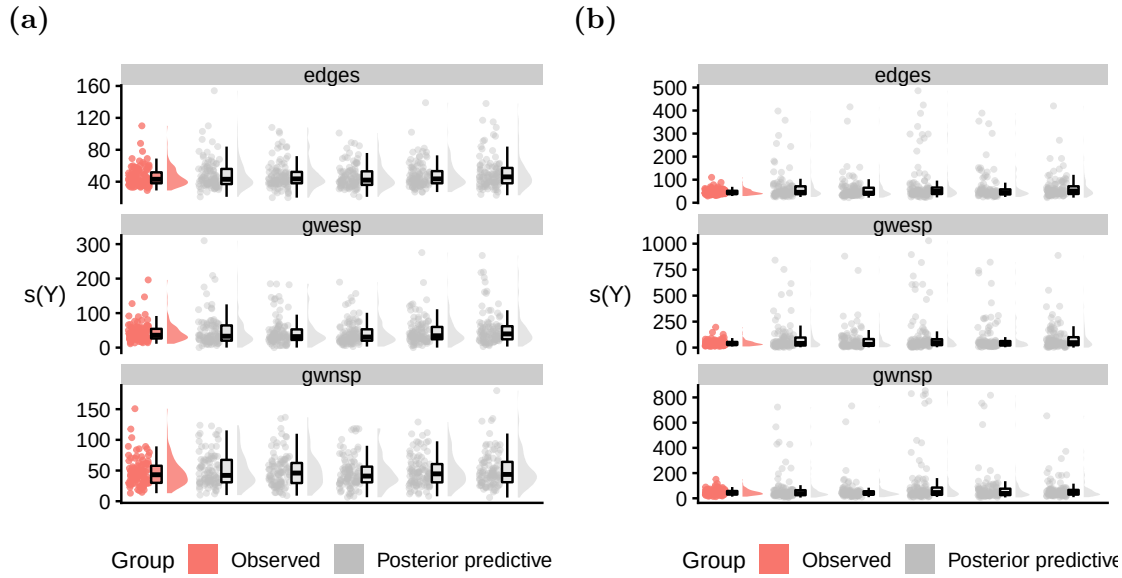


Figure 2.16: Posterior predictive checks under the vague prior. (a) When simulating populations of networks based on individual-level resampling, the model fit (in terms of the model summary statistics) appears to be good with the simulated populations closely matching the observed population. (b) However, under group-level resampling, the simulated populations generally have a wider range of summary statistic values compared to the observed population, suggesting that the covariance $\mathbf{\Sigma}_\theta$ is poorly estimated.

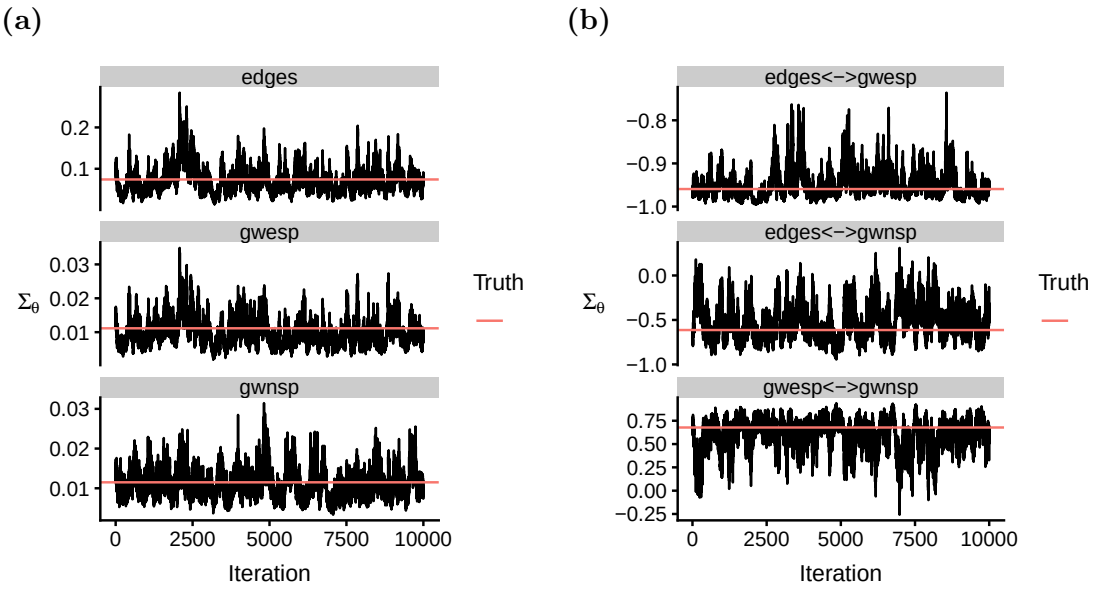


Figure 2.17: Under the empirical prior, the sequence of $\mathbf{\Sigma}_\theta$ samples covers the underlying truth for both (a) the diagonal elements of $\mathbf{\Sigma}_\theta$ and (b) the correlation of the $\theta^{(i)}$ parameters, i.e. the normalised off-diagonal elements of $\mathbf{\Sigma}_\theta$.

(Figure 2.16b). This is due to the fact that $\mathbf{\Sigma}_\theta$ is not being captured accurately. In particular, the sequence of samples for off-diagonal elements of $\mathbf{\Sigma}_\theta$ underestimate the strong correlation within the $\theta^{(i)}$ parameters. Thus, some of the simulated networks based on new values of $\theta^{(i)}$ have summary statistics that are vastly different to those in the observed population.

While increasing the number of networks n in the population would likely lead to more accurate estimates, we propose an alternative approach to improve estimation of $\mathbf{\Sigma}_\theta$. Using the individual Bayesian ERGM fits described in Section 2.5.2, we construct an empirical prior for $\mathbf{\Sigma}_\theta$. With estimates $\bar{\theta}^{(1)}, \ldots, \bar{\theta}^{(n)}$ of the posterior mean from the individual fits and $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} \bar{\theta}^{(i)}$, we set the prior of $\mathbf{\Sigma}_\theta$ to be:

$$\mathbf{\Sigma}_\theta \sim \mathcal{W}^{-1}(\hat{\mathbf{\Sigma}}_\theta, p+1) \tag{2.34}$$

where $\hat{\mathbf{\Sigma}}_\theta = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{\theta}^{(i)} - \bar{\mu})(\bar{\theta}^{(i)} - \bar{\mu})^T$. When using this empirical prior, the model fitting procedure (Algorithm 2.5) produces more accurate samples for $\mathbf{\Sigma}_\theta$ (Figure 2.17). This is also reflected in the posterior predictive checks: the simulated populations of networks have similar summary statistics to those of the observed population under individual-level resampling (Figure 2.18a) and group-level resampling (Figure 2.18b). This highlights the

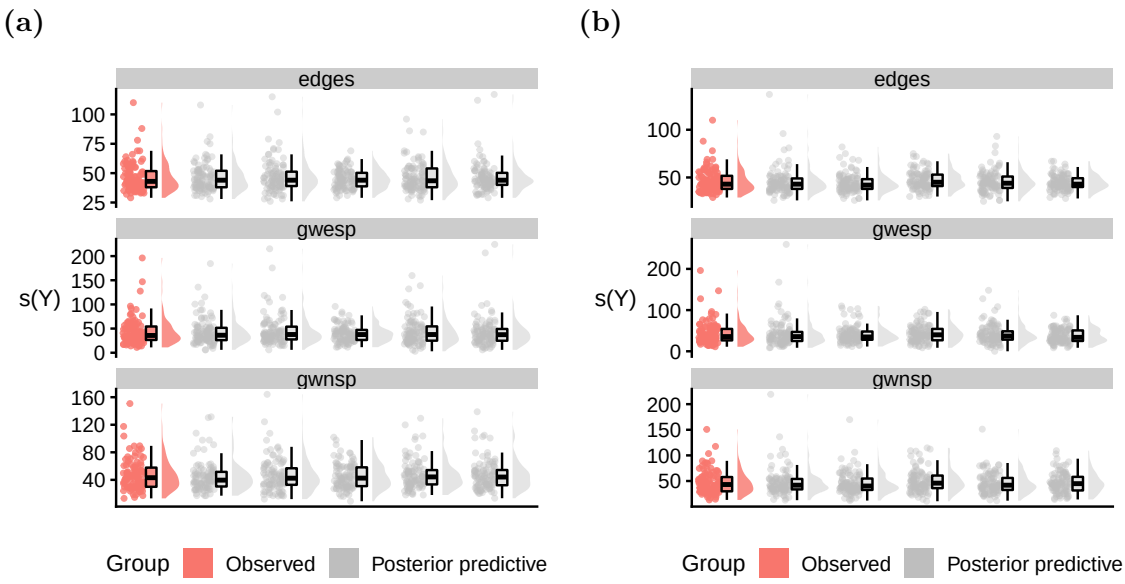


Figure 2.18: Posterior predictive checks under the empirical prior. The model fit (in terms of the model summary statistics) appears to be good when simulating populations of networks via both (a) individual-level resampling, and (b) group-level resampling.

pragmatic need of the empirical prior - with small n, we are unlikely to learn $\boldsymbol{\Sigma}_\theta$ from an uninformative prior.

2.5.5 Further adaptations

The adaptations described in this section were not included in the final fitting procedure for reasons discussed below but were nonetheless investigated in pursuit of improving computational efficiency. We first introduce the concepts of parameter expansion (see e.g. [92]) and blockwise sampling (see e.g. [161]), both common techniques to improve the efficiency of Bayesian hierarchical models. We then discuss an idea to reduce the number of ERG simulations in each MCMC iteration, based on a Taylor expansion of the acceptance ratio.

Parameter expansion, in the context of Bayesian hierarchical models, introduces a new parameter $\zeta = diag(\zeta_1, \ldots, \zeta_p) \in \mathbb{R}^{p \times p}$ into the model, in order to break the dependence between the individual-level parameters $\tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(n)}$ and the group-level covariance

parameter $\mathbf{\Sigma}_\theta$. In the non-centred parametrisation, we write the model as:

$$\begin{aligned} Y^{(i)} &\sim ERGM(\mu + \zeta\tilde{\theta}^{(i)}), \quad i = 1, \ldots, n \\ \tilde{\theta}^{(i)} &\sim \mathcal{N}(0, \mathbf{\Sigma}_\theta), \quad i = 1, \ldots, n \\ \mu &\sim \mathcal{N}(\alpha, \mathbf{\Sigma}_\mu) \\ \zeta_i &\sim U(-\infty, \infty) \\ \mathbf{\Sigma}_\theta &\sim \mathcal{W}^{-1}(\nu, \mathbf{\Psi}) \end{aligned} \tag{2.35}$$

We can adapt the model fitting algorithms by introducing an extra exchange step to update the ζ_i parameters. The parameter expansion strategy is useful when the estimates for $\mathbf{\Sigma}_\theta$ are close to zero. This can cause the algorithm to get stuck, as the $\tilde{\theta}^{(i)}$ also shrink towards zero, and then the next update of $\mathbf{\Sigma}_\theta$ is likely to be close to zero again. Introducing the ζ_i parameters breaks this dependence and allows the algorithm to escape this behaviour. We find however, that in practice Algorithm 2.5 does not exhibit this behaviour and so, given the extra computational cost, choose not to include the extra ζ parameter in the model.

Block-wise sampling consists of updating batches of model parameters together in one step. In Algorithm 2.5, each of the parameters $\mu, \mathbf{\Sigma}_\theta, \tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(n)}$ are updated one-at-a-time. There are of course various ways to combine the different model parameters into batches. One intuitive way to implement block-wise sampling is to partition the individual-level parameters $\tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(n)}$ into g groups where $g < n$. To update the j^{th} block $\boldsymbol{\theta}^j = (\tilde{\theta}^{(i_1)}, \ldots, \tilde{\theta}^{(i_{n_j})})$: draw a proposal $\boldsymbol{\theta}'$, simulate exponential random graphs $\mathbf{y}^{(k)'} \sim ERGM(\mu + \tilde{\theta}^{(i_k)})$ for $k = 1, \ldots, n_j$, and accept the proposal with probability

$$\min\left(1, \frac{\exp\left\{\sum_{k=1}^{n_j} \tilde{\theta}^{(i_k)'T}[s(\mathbf{y}^{(i_k)}) - s(\mathbf{y}^{(i_k)'})] - \frac{1}{2}\tilde{\theta}^{(i_k)'T}\mathbf{\Sigma}_\theta^{-1}\tilde{\theta}^{(i_k)'}\right\}}{\exp\left\{\sum_{k=1}^{n_j} \tilde{\theta}^{(i_k)T}[s(\mathbf{y}^{(i_k)}) - s(\mathbf{y}^{(i_k)'})] - \frac{1}{2}\tilde{\theta}^{(i_k)T}\mathbf{\Sigma}_\theta^{-1}\tilde{\theta}^{(i_k)}\right\}}\right). \tag{2.36}$$

We found, however, that using block-wise sampling did not significantly improve the computational efficiency of the algorithm and so we only present results using component-wise sampling.

As mentioned previously, the main computational bottleneck in the model fitting procedure lies in the exchange step, due to the need to simulate a population of exponential random graphs. Ideally, we would reduce the number of network simulations at each MCMC iteration. Here we discuss one possible approach towards this goal, centred

around the concept of *noisy Monte Carlo* [9].

The key idea behind noisy Monte Carlo, in the context of the Metropolis Algorithm 2.1 for Bayesian ERGMs, is to replace the acceptance ratio $AR(\theta', \theta_{k-1}; \mathbf{y})$ by an estimator $\hat{AR}(\theta', \theta_{k-1}; \mathbf{y})$. Alquier et al. [9] prove that if $|AR(\theta', \theta_{k-1}; \mathbf{y}) - \hat{AR}(\theta', \theta_{k-1}; \mathbf{y})|$ is small enough (along with certain other conditions), then using the estimator yields a Markov chain that approximates the original target distribution.

Recall that the obstacle in the Metropolis algorithm is due to the ratio of intractable normalising constants $\frac{Z(\theta_{k-1})}{Z(\theta')}$ in the acceptance ratio:

$$AR(\theta', \theta_{k-1}; \mathbf{y}) = \frac{\exp\left\{\theta'^T s(\mathbf{y})\right\} \pi(\theta')}{\exp\left\{\theta_{k-1}^T s(\mathbf{y})\right\} \pi(\theta_{k-1})} \cdot \frac{Z(\theta_{k-1})}{Z(\theta')}. \tag{2.37}$$

Since we can evaluate $\dfrac{\exp\left\{\theta'^T s(\mathbf{y})\right\} \pi(\theta')}{\exp\left\{\theta_{k-1}^T s(\mathbf{y})\right\} \pi(\theta_{k-1})}$ directly, we need only construct an estimator for $\dfrac{Z(\theta_{k-1})}{Z(\theta')}$. To do so, we will write $\epsilon = \theta' - \theta_{k-1}$ and begin with a 2^{nd} order Taylor expansion of the *logarithm* of $Z(\theta_{k-1})$ around θ':

$$\begin{aligned} \log \frac{Z(\theta_{k-1})}{Z(\theta')} &= \log Z(\theta' + \epsilon) - \log Z(\theta') \\ &\approx \epsilon^T \nabla \log Z(\theta') + \frac{1}{2} \epsilon^T \left(\nabla^2 \log Z(\theta')\right) \epsilon \end{aligned} \tag{2.38}$$

At first sight, this approximation does not appear to be particularly useful; $Z(\theta')$ is intractable so both $\nabla \log Z(\theta')$ amd $\nabla^2 \log Z(\theta')$ are also intractable. However, we can use the properties of exponential family distributions to construct estimators to these terms. First, note that, we can write the moment generating function of the sufficient statistics of $\mathbf{Y}$:

$$\begin{aligned} M(t) :&= \mathbb{E}_\theta[\exp(t^T s(\mathbf{y}))] \\ &= \sum_{\mathbf{y}' \in \mathcal{Y}} \exp(t^T s(\mathbf{y}')) \cdot \frac{\exp(\theta^T s(\mathbf{y}'))}{Z(\theta)} \\ &= \frac{1}{Z(\theta)} \sum_{\mathbf{y}' \in \mathcal{Y}} \exp((t + \theta)^T s(\mathbf{y}')) \\ &= \frac{Z(t + \theta)}{Z(\theta)} \end{aligned} \tag{2.39}$$

Thus, the cumulant generating function of the sufficient statistics is:

$$\begin{aligned} K(t) :&= \log M(t) \\ &= \log Z(t+\theta) - \log Z(\theta) \end{aligned} \tag{2.40}$$

We can use this fact to write the cumulants of the sufficient statistics of $\mathbf{Y}$ in terms of the derivatives of $\log Z(\theta)$. In particular,

$$\mathbb{E}_\theta[s(\mathbf{Y})] = \nabla \log Z(\theta) \tag{2.41}$$

and

$$\mathrm{Cov}_\theta[s(\mathbf{Y})] = \nabla^2 \log Z(\theta). \tag{2.42}$$

While $\mathbb{E}_\theta[s(\mathbf{Y})]$ and $\mathrm{Cov}_\theta[s(\mathbf{Y})]$ are generally not available in closed form, we may estimate them using Monte Carlo integration. $\mathbb{E}_\theta[s(\mathbf{Y})]$ can be approximated by $M_{N,\theta} = \frac{1}{N}\sum_{i=1}^{N} s(\mathbf{y}^i)$ where $\mathbf{y}^i \sim ERGM(\theta)$ for $i = 1, \ldots, N$. Similarly, we can approximate $\mathrm{Cov}_\theta[s(\mathbf{Y})]$ by the sample variance $\mathbf{S}_{N,\theta} = \frac{1}{N-1}\sum_{i=1}^{N}(s(\mathbf{y}^i) - M_{N,\theta})(s(\mathbf{y}^i) - M_{N,\theta})^T$.

Let us return to the Taylor approximation given in (2.38). Combining this with the above Monte Carlo approximations, we have

$$\log \frac{Z(\theta_{k-1})}{Z(\theta')} \approx \epsilon^T M_{N,\theta'} + \frac{1}{2}\epsilon^T \mathbf{S}_{N,\theta'}\epsilon. \tag{2.43}$$

For a single Bayesian ERGM exchange algorithm step, this approximation does not yield any computational gains, as we must simulate $N \geq 2$ exponential random graphs to construct $M_{N,\theta'}$ and $\mathbf{S}_{N,\theta'}$. However, for the multilevel model, this approach may be computationally cheaper if we choose N to be less than the number of networks n.

To see this, observe that at the k^{th} iteration we may write the individual-level parameters and their respective proposals as $\theta_{k-1}^{(i)} = \mu_k + \epsilon_{k-1}^{(i)}$ and $\theta^{(i)'} = \mu_k + \epsilon^{(i)'}$ for $i = 1, \ldots, n$. Then, using the Taylor approximation for both the denominator and numerator of $\log \dfrac{Z(\theta_{k-1}^{(i)})}{Z(\theta^{(i)'})}$, we have

$$\log \frac{Z(\theta_{k-1}^{(i)})}{Z(\theta^{(i)'})} \approx (\epsilon_{k-1}^{(i)} - \epsilon^{(i)'})^T M_{N,\mu_k} + \frac{1}{2}\epsilon_{k-1}^{(i)T}\mathbf{S}_{N,\mu_k}\epsilon_{k-1}^{(i)} - \frac{1}{2}\epsilon^{(i)'T}\mathbf{S}_{N,\mu_k}\epsilon^{(i)'}, \quad i = 1, \ldots, n. \tag{2.44}$$

Thus, to update the n individual-level parameters, we need only simulate N exponential

random graphs, where N may be less than n.

This gain in computational efficiency does come at the cost of two levels of approximation: one in the Taylor series expansion, and the second in the Monte Carlo integration. In both cases, we are able to control the amount of error introduced. The Taylor approximation error can be reduced by using smaller proposal jumps. In using smaller jumps, however, the autocorrelation of the sequence of samples is likely to be higher, thus reducing the effective sample size and potentially cancelling any gains in efficiency. Similarly, the Monte Carlo error can be reduced by increasing the number of Monte Carlo simulations N. Yet, choosing N to be larger than n defeats the point of using this approach in the first place. To understand when the approach would be effective, we would need to quantify the error for both levels of approximation. Although this warrants further investigation, we focus instead on the development of the general framework

2.6 Future work

The main contribution of this chapter was to propose a multilevel framework for modelling populations of networks, along with a novel MCMC procedure for performing inference with the framework. While the framework itself is a natural multilevel extension of a single ERGM, the development of the associated inference procedure proved challenging. To reduce the computational cost of the fitting procedure, we explored a number of adaptations, not all of which were included in the final algorithm. Future work should focus on further improving computational efficiency in order to make the framework more practical and accessible to use. For example, as discussed in the previous section, it may be possible to reduce the number of ERGM simulations at each MCMC iteration using noisy Monte Carlo methods [9]. Another promising avenue is variational inference for ERGMs [144], which could be extended to our framework to yield approximate Bayesian inference at a much reduced computational cost relative to MCMC.

Chapter 3

Inferring group-level differences in functional connectivity with Bayesian exponential random graph models

Summary

The goal of many neuroimaging studies is to better understand how the functional connectivity structure of the brain changes with a given phenotype such as age. Functional connectivity is commonly characterised as a network, with nodes corresponding to brain regions and edges corresponding to statistical dependencies between the respective regional time series of activity. To assess group-level variations in functional connectivity, it is therefore crucial to develop statistical methods for the inference of differences in the topological structure between groups of networks. We propose one such approach based on the multilevel framework introduced in the previous chapter. Our motivation stems from the Cam-CAN project, a neuroimaging study on healthy ageing. Using this dataset, we illustrate how our method can be used to infer differences in functional connectivity between a group of young individuals and a group of old individuals.

The chapter consists of the following: Section 3.1 reviews some of the existing approaches to brain connectivity comparisons; Section 3.2 introduces the data to which we will apply our method; Section 3.3 describes how to extend the multilevel framework to perform group comparisons; Section 3.4 reports the results of applying our method to the

Cam-CAN data; and Section 3.5 discusses the implications and limitations of our findings.

3.1 Background

Network analyses have become a popular tool for relating brain connectivity to a variety of phenotypes such as neurological disease [5, 32], cognitive function [155], and age [52]. The broad aim of such studies is to understand how the brain connectivity differs between individuals in distinct groups, for example, young versus old. Given the complex structure of network data, this is a challenging task and the development of statistical methods to infer differences between networks is still in its infancy. The focus of this chapter is on brain functional connectivity, though the ideas presented here are applicable to the general task of comparing groups of networks.

Notation

Before reviewing existing approaches for comparing groups of networks, we first recap and introduce some notation. Denote J to be the number of groups, and the number of individuals in group j as n_j (allowing for unbalanced groups, $n_1 \neq n_2$ and assuming that the group memberships are known). We assume that the set $\mathcal{V}$ of nodes is fixed and common among all networks in all groups, and let $V = |\mathcal{V}|$ be the number of nodes. We will work directly with the adjacency matrices associated with each individual's network. Let $\mathbf{Y}^{(i,j)}$ be the (random) adjacency matrix associated with the i^{th} individual in group j, and denote $\mathbf{y}^{(i,j)}$ to be an instantiation of $\mathbf{Y}^{(i,j)}$. Thus, $\mathbf{Y}^{(i,j)} \in \{0,1\}^{V \times V}$ with $Y_{kl}^{(i,j)}$ indicating a connection between regions k and l. We assume that the range of the $\mathbf{Y}^{(i,j)}$, i.e. the set of all possible outcomes, is common across all networks and denote this $\mathcal{Y}$. Let $s(\mathbf{Y}^{(i,j)})$ denote a vector of p summary statistics of $\mathbf{Y}^{(i,j)}$, such that each component is a function $s_i : \mathcal{Y} \to \mathbb{R}$. We write $\boldsymbol{y}^{(j)} = (\mathbf{y}^{(1,j)}, \ldots, \mathbf{y}^{(n_j,j)})$ for the set of networks in group j, and $\boldsymbol{y} = (\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})$ for the set of all networks.

Existing approaches

One possible approach to comparing brain networks across groups is to test for differences across all edges separately. The main drawback of this mass-univariate approach is the multiple testing issue: the significance threshold must be adjusted for the $\binom{V}{2}$ comparisons being made simultaneously. As the number of multiple comparisons being made is

relatively large, using a standard false discovery rate (FDR) correction may result in a test with insufficient power. Scott et al. [126] introduce edge covariate information (such as spatial or functional relationships) to improve power. Zalesky et al. [168] proposed an alternative thresholding procedure that instead compares connected components between groups using a permutation testing strategy. While these approaches are useful for testing for differences in individual connections, or sets of connections, between groups, they are not able to detect variations in the global network structure.

Another family of methods proceeds by reducing each network $\mathbf{Y}^{(i,j)}$ to a set of p summary statistics $s(\mathbf{Y}^{(i,j)})$, typically chosen to reflect some global network properties of interest. One approach is then to perform a standard (multivariate) regression procedure to test for a group difference on the summary statistics (e.g. [35]). Another popular approach is permutation testing (e.g. [133]). Permutation testing compares the observed difference in summary statistics between groups to a null distribution generated by repeatedly reallocating the group labels of the data.

One major drawback of these approaches is that they do not account for the relationship between summary statistics arising due to the structure of network. For example, a network with more edges typically has more triangles. This issue has previously been studied by Ginestet et al. [60]. They show, in particular, that a variety of popular summary statistics, such as global efficiency and network modularity, are sensitive to changes in the number of edges in the network. Therefore, group differences in these more complex network metrics may simply reflect differences in the overall density of the network. This is particularly important in the context of neuroimaging studies: differences in connectivity structure may arise due to other factors of the data, such as differing levels of noise.

Another approach based on summary statistics uses exponential random graph models (ERGMs; see Section 2.1). Simpson et al. [133] used ERGMs to construct group-based representative connectivity networks. Recently, there has been renewed interest in using such statistical network models for brain connectivity comparisons. Sinke et al. [134] applied a Bayesian ERGM to compare structural connectivity networks. Ginestet et al. [61] used graph Laplacians to construct hypothesis tests of differences in network data. Durante & Dunson [41] proposed a Bayesian nonparametric model that allows for the testing of both local (edge-based) and global (whole network) differences.

While these developments are an important and welcome contribution towards the goal of comparing brain networks, they typically assume that the networks for individuals

in the same group are drawn from a common distribution. Since functional connectivity networks have been shown to be highly variable across individuals [101], there is a need for models that are able to capture the variability as well as the commonalities in connectivity structure between individuals in the same group.

Motivated by this issue, we show how the multilevel framework presented in Chapter 2 can be used to jointly model individual functional connectivity networks. We demonstrate our method using resting-state fMRI data from the Cam-CAN study [127], described in Section 3.2. In Section 3.3, we extend the framework to deal with multiple groups of individuals, and show how one can use the framework to detect group-level differences in connectivity structure. Section 3.4 presents the results from applying our framework to the Cam-CAN data, and Section 3.5 discusses the limitations of our approach.

3.2 Data

The functional scans were collected both in resting-state and a variety of task-based experiments. Of these, we used the resting-state fMRI data to construct functional connectivity networks. To assess how the functional connectivity structure differs between young individuals and old individuals in the Cam-CAN dataset, we focused on the youngest (18-27, 52 subjects) and oldest (78-88, 95 subjects) decades. Section 3.2.1 outlines the preprocessing steps undertaken to prepare the data for analysis. Section 3.2.2 describes how we construct individual networks from the time series data.

3.2.1 Preprocessing

The data were collected as part of Phase II of the CamCAN project (www.cam-can.org; [127]). The MRI data were acquired on a 3T Siemens TIM Trio at the MRC Cognition & Brain Sciences Unit, with a 32 channel head-coil. Structural images were acquired using a 1mm3 isotropic, T1-weighted Magnetization Prepared RApid Gradient Echo (MPRAGE) sequence and a 1mm3 isotropic, T2-weighted Sampling Perfection with Application optimized Contrasts using different flip angle Evolution (SPACE) sequence. The fMRI data for the eyes-closed, resting-state run were acquired using a Gradient-Echo (GE) Echo-Planar Imaging (EPI) sequence, and consisted of 261 volumes (lasting 8 min and 40 s). Each volume contained 32 axial slices (acquired in descending order), with slice thickness of 3.7 mm and interslice gap of 20% (for whole brain coverage including cere-

bellum; TR 1,970 ms; TE 30ms; voxel-size 3 mm x 3 mm x 4.44 mm). EPI fieldmaps with two TEs (5.19ms and 7.65ms) were also acquired. All the raw data from Cam-CAN Phase II, together with more acquisition details, are available on: http://www.cam-can.org/index.php?content=dataset.

The data were processed using r7219 of the SPM12 software (http://www.fil.ion.ucl.ac.uk/spm), automated with r5.4 of the Automatic Analysis (AA) pipeline system [33] (http://github.com/automaticanalysis/automaticanalysis/releases/v5.4.0; for overview of pipelines, see [145]) in r2015 of MATLAB (The MathWorks). The current data come from local Release0004 of the CamCAN pipelines. To obtain a good starting-point for image normalisation, the T1 image was coregistered to the Montreal Neurological Institute (MNI) template using rigid body transformation, and then the T2 image then coregistered to the T1. Both T1 and T2 images were bias corrected, and then combined in a multimodal segmentation to estimate images of each of six tissue classes, including grey-matter (GM), white-matter (WM) and cerebrospinal fluid (CSF). Diffeomorphic registration (DARTEL) was then applied to the GM and WM segments to create group templates, which were in turn transformed to MNI space using a 12-parameter affine transform.

The fMRI images were unwarped using distortion fields estimated from the fieldmaps, and corrected for motion using rigid-body realignment to the mean fMRI image across runs. The different slice acquisition times were corrected by interpolating to the middle slice. The images were rigid-body coregistered to the T1 image and the spatial transformations from that T1 to MNI space (diffeomorphic and affine) applied to every fMRI image. Residual effects of abrupt motion were reduced by applying wavelet despiking [108]. The mean time series for all voxels within the thresholded WM and CSF segments were calculated to use as later covariates of no interest.

The fMRI time series were then extracted from 38 regions of interest (ROIs) from the OSL toolbox (https://github.com/OHBA-analysis/parcellations). The time series for each voxel in each ROI were adjusted for various confounds by taking the residuals from a general linear model (GLM) that included: 1) the time series in WM and CSF segments, 2) the 6 rigid-body motion parameters from the realignment stage, 3) the first-order difference in those motion parameters across successive TRs and 4) a cosine basis set up to cut-off frequency of 0.008Hz (implementing a form of high-pass filter). Second-order (squared) terms for 1), 2) and 3) were also included in the GLM. The autocorrelation in the GLM error was modelled by a family of 8 exponentials with half-lives from 0.5 to 64

TRs (estimated by pooling across voxels within each ROI), and used to prewhiten the time series. The first eigenvector across voxels of the residual time series was then used as a summary measure for each ROI. This approach was based on the optimised pipeline proposed by Geerligs et al. [54].

3.2.2 Network construction

The preprocessed data consists of $N = 38$ ROI time series for each individual. To extract a connectivity network for a given individual i, we first computed the pairwise Pearson correlation between each of the time series, yielding a $N \times N$ correlation matrix $\mathbf{C}^{(i)}$. We then applied a threshold $r^{(i)}$ to each correlation matrix to produce an $N \times N$ adjacency matrix $\mathbf{y}^{(i)}$ with entries:

$$\mathbf{y}_{kl}^{(i)} = \begin{cases} 1 & \text{if } \mathbf{C}_{kl}^{(i)} \geq r^{(i)} \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

The choice of threshold $r^{(i)}$ can have a significant impact on the resulting connectivity structure [156, 152]. The two most common strategies are "absolute" thresholding and "proportional" thresholding. Absolute thresholding picks a common threshold $r^{(i)} = r$ for each individual. In contrast, proportional thresholding ensures that the same number of edges are present in each network by allowing the threshold to vary by individual.

Both thresholding strategies have their drawbacks. Since the number of edges in a network inherently affects the overall network [156, 60], differences in other network metrics may simply be attributable to variations in the overall connectivity. A proportional threshold may thus be preferred in order to keep the number of edges constant across individuals (in both groups) and facilitate comparison of other metrics of interest. On the other hand, a lower correlation value may be less reliable in indicating a functional connection between brain regions. Therefore, by including lower correlations as edges for individuals with lower overall connectivity, a proportional threshold may induce more randomness in the resulting network [152]. This is important in the context of group studies for ageing: overall functional connectivity tends to decrease with age (Figure 3.1). Given these issues, we present our findings using the two different thresholding procedures (absolute and proportional) at two different thresholds (10% and 20%).

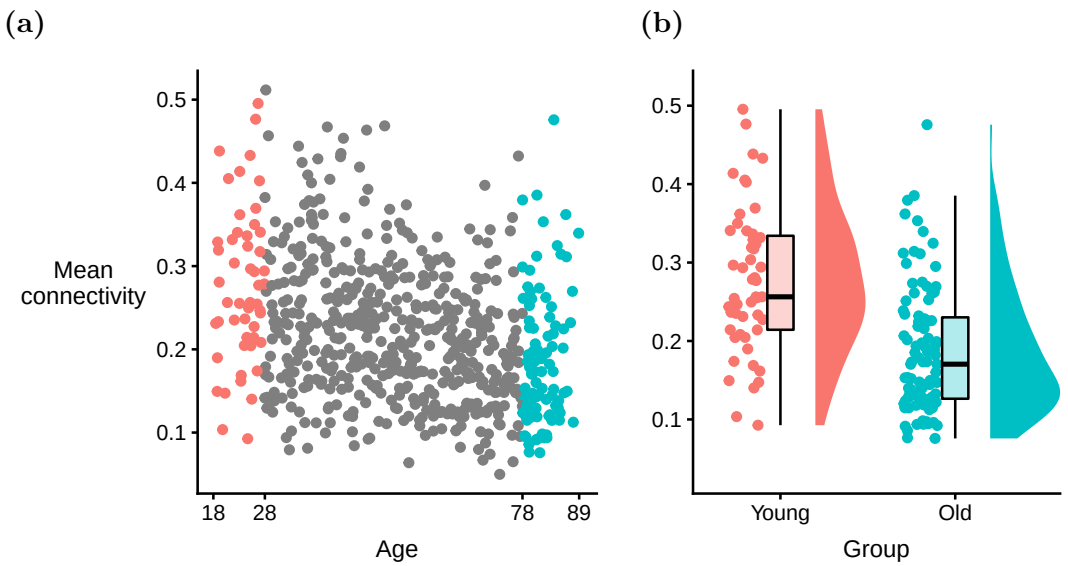


Figure 3.1: Mean connectivity against age

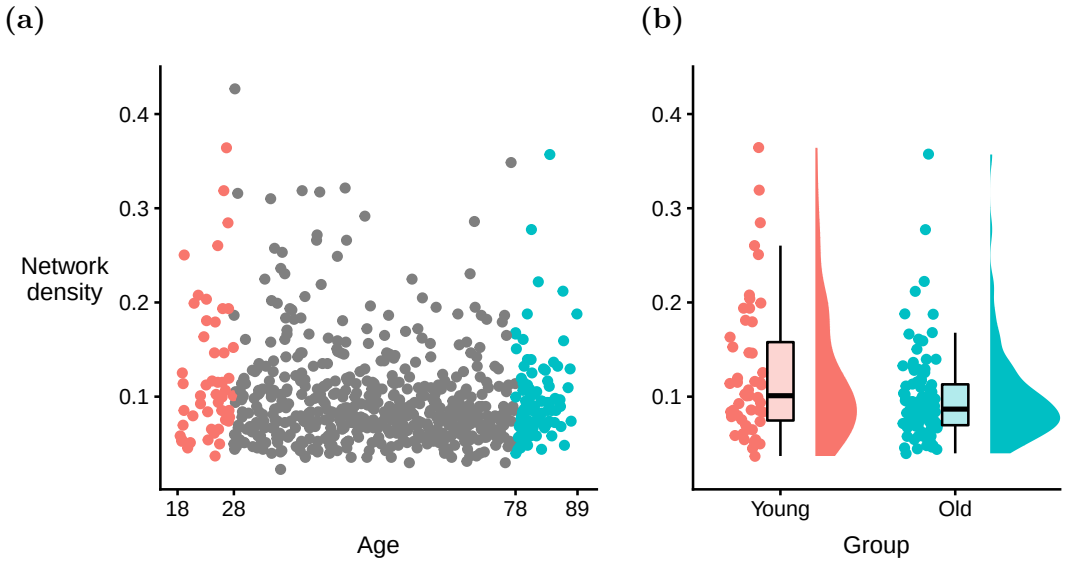


Figure 3.2: Network density against age, constructed with 10% absolute threshold.

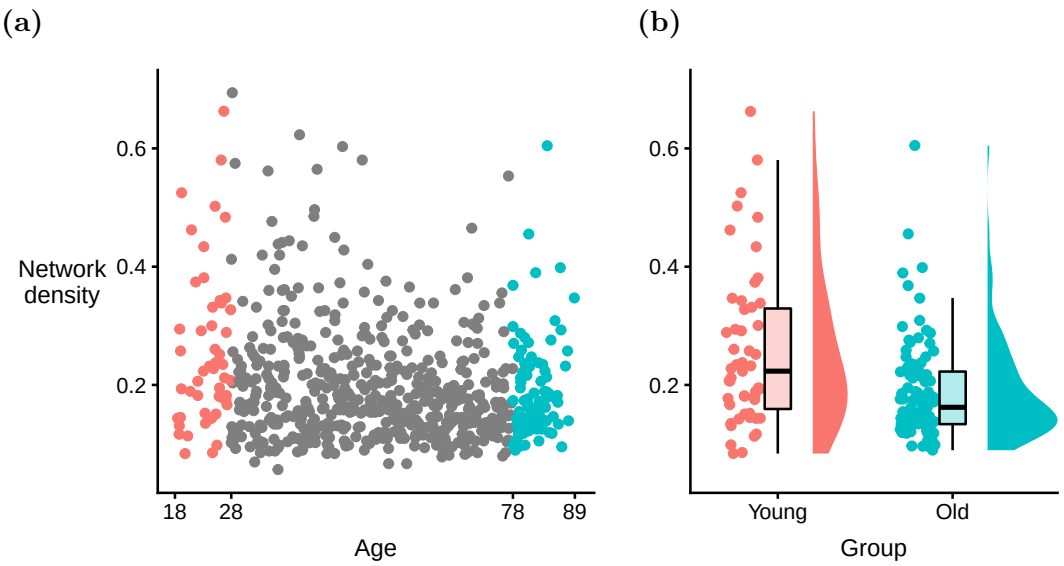


Figure 3.3: Network density against age, constructed with 20% absolute threshold.

3.3 Methods

In Chapter 2, we proposed a multilevel framework to model a population of networks. To briefly recap, this framework models each network $\mathbf{Y}^{(i)}$ as an exponential random graph with individual-level parameter $\theta^{(i)}$. Next, we assume that the individual-level parameters are draw from a group-level distribution $\theta^{(i)} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Here, we describe in detail how to extend this framework to deal with group structure within the population. Section 3.3.1 fully specifies the model, including the priors; and Section 3.3.2 details how the framework can be used to detect group differences.

3.3.1 Model specification

In Section 2.3, we briefly outlined how the multilevel framework could be extended to deal with group structure in the data by adding a level of hierarchy to the model. To investigate possible group differences, we discuss three separate ways to model the group structure of the individuals, corresponding to 1) no pooling, 2) complete pooling, and 3) partial pooling.

ERGM summary statistics

For the absolute thresholding, we modelled each individual network as an exponential random graph with three summary statistics that have previously been applied successfully to fit single brain functional connectivity networks [132, 131]. The three summary statistics used are 1) number of edges, 2) geometrically weighted edgewise shared partners (GWESP), and 3) geometrically weighted non-edgewise shared partners (GWNSP). For the proportional thresholding, we fixed the number of edges and modelled the network using two summary statistics: GWESP and GWNSP.

The number of edges $E = \sum_{i<j} Y_{ij}$ characterises the sparsity of the network. Mean functional connectivity decreases with age (Figure 3.1) so, under absolute thresholding, networks for the old group tend to have fewer edges than the young group (Figures 3.2 and 3.3). It is important to include this summary statistic in the model in order to account for this variation between groups.

The geometrically weighted edgewise shared partner (GWESP) statistic of a network $\mathbf{y}$ is given by:

$$GWESP(\mathbf{y}) = e^{\tau_1} \sum_{i=1}^{V-2} \left\{1 - (1 - e^{-\tau_1})^i\right\} EP_i(\mathbf{y}) \tag{3.2}$$

where $EP_i(\mathbf{y})$ is the number of *connected* node pairs having exactly i shared partners and $\tau_1 > 0$ is a decay parameter. Note that the $EP_i(\mathbf{y})$ are closely related to the number of k-triangles $T_k(\mathbf{y})$ through the equations:

$$\begin{aligned} T_1(\mathbf{y}) &= \frac{1}{3} \sum_{i=1}^{V-2} iEP_i(\mathbf{y}) \\ T_k(\mathbf{y}) &= \sum_{i=k}^{V-2} \binom{i}{k} EP_i(\mathbf{y}), \quad k = 2, \ldots, V-2 \end{aligned} \tag{3.3}$$

Thus, the GWESP statistic can be seen a measure of network transitivity (or clustering) that accounts not only for the number of 1-triangles but also the number of k-triangles for $k > 1$. The decay parameter τ_1 serves to diminish the effect of the network having more higher-order edgewise shared partners relative to lower-order edgewise shared partners. In other words, the increase in GWESP from adding a single edge is smaller if the edge adds a shared partner to connected nodes that already share many partners.

The geometrically weighted non-edgewise shared partner (GWNSP) statistic is simi-

larly defined as

$$GWNSP(\mathbf{y}) = e^{\tau_2} \sum_{i=1}^{V-2} \left\{1 - (1 - e^{-\tau_2})^i\right\} NP_i(\mathbf{y}) \tag{3.4}$$

where $NP_i(\mathbf{y})$ is the number of *non-connected* node pairs having exactly i shared partners and $\tau_2 > 0$ is a decay parameter. GWNSP is related to global network efficiency; a higher value of GWNSP indicates that non-connected nodes are more likely to have a shared partner.

While it is possible to treat the decay parameters for both geometrically weighted statistics as extra model parameters (leading to *curved* ERGMs [75]), this increases the computational burden substantially. The decay parameters were therefore fixed at $\tau_1 = \tau_2 = 0.75$, as these values generally resulted in better fitting models [132].

We write the probability mass function of the network of the i^{th} individual in group j as

$$\pi(\mathbf{y}^{(i,j)}|\theta^{(i,j)}) = \frac{\exp\{\theta^{(i,j)T} s(\mathbf{y}^{(i,j)})\}}{Z(\theta^{(i,j)})} \tag{3.5}$$

where $\theta^{(i,j)}$ is a vector of individual-level model parameters, and $s(\mathbf{y}^{(i,j)})$ is the vector of observed summary statistics. Note that, under proportional thresholding, the sample space $\mathcal{Y}$ is reduced to networks with the same number of edges as those observed. For example, with a 10% threshold, each constructed network has $E = 70$ edges, so the sample space is all possible networks with 70 edges.

Group structure

The three approaches to modelling the group structure differ only in the specification of the priors for the individual-level parameters. The first, *no pooling* approach assumes that the two groups are completely independent. In particular, we assume that the individual-level parameters for the two groups are drawn from distinct group-level distributions:

$$\theta^{(i,j)} \sim \mathcal{N}(\mu^{(j)}, \mathbf{\Sigma}_\theta^{(j)}), \quad j = 1, 2, \quad i = 1, \ldots, n_j. \tag{3.6}$$

In turn, we specify conditionally conjugate priors for the group-level parameters $(\mu^{(j)}, \mathbf{\Sigma}_\theta^{(j)})$:

$$\begin{aligned} \mu^{(j)} &\sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_\mu), \ j = 1, 2 \\ \mathbf{\Sigma}_\theta^{(j)} &\sim \mathcal{W}^{-1}(\mathbf{\Psi}_\theta, \nu_\theta), \ j = 1, 2 \end{aligned} \tag{3.7}$$

The second, *complete pooling* approach assumes that all individuals belong to the same group, thus ignoring any potential group differences between the young and the old individuals. Specifically, we assume:

$$\theta^{(i,j)} \sim \mathcal{N}(\mu, \mathbf{\Sigma}_\theta), \quad j = 1, 2, \quad i = 1, \ldots, n_j. \tag{3.8}$$

We specify the same conditionally conjugate priors on the group-level parameters $(\mu, \mathbf{\Sigma}_\theta)$:

$$\begin{aligned} \mu &\sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_\mu) \\ \mathbf{\Sigma}_\theta &\sim \mathcal{W}^{-1}(\mathbf{\Psi}_\theta, \nu_\theta) \end{aligned} \tag{3.9}$$

The third, *partial pooling*, approach sits somewhere between the previous two. We allow the individual-level parameters to be drawn from two normal distributions with different group-level means, but assume that the covariance structure is common across the groups:

$$\theta^{(i,j)} \sim \mathcal{N}(\mu^{(j)}, \mathbf{\Sigma}_\theta), \quad j = 1, 2, \quad i = 1, \ldots, n_j. \tag{3.10}$$

In contrast to the first two approaches, we change the specification of the group-level hyperprior distributions. Most importantly, we add another level of hierarchy for the group-level means:

$$\begin{aligned} \mu^{(j)} &\sim \mathcal{N}(\mu^{pop}, \mathbf{\Sigma}_\mu) \\ \mu^{pop} &\sim \mathcal{N}(\mu_0, \mathbf{\Lambda}) \\ \mathbf{\Sigma}_\mu &\sim \mathcal{W}^{-1}(\nu_\mu, \mathbf{\Psi}_\mu) \end{aligned} \tag{3.11}$$

This allows information to be pooled across the two groups via the population-level hyperparameter μ^{pop} while also allowing for differences in the group-level means. For the group-level covariance $\mathbf{\Sigma}_\theta$, we specify an inverse-Wishart hyperprior:

$$\mathbf{\Sigma}_\theta \sim \mathcal{W}^{-1}(\nu_\theta, \mathbf{\Psi}_\theta) \tag{3.12}$$

Given the additional parameters, we need to adapt Algorithm 2.5 in order to produce samples from the joint posterior $\pi(\boldsymbol{\theta}, \mu^{(1)}, \mu^{(2)}, \mathbf{\Sigma}_\theta, \mu^{pop}, \mathbf{\Sigma}_\mu | \boldsymbol{y})$. Due to the (partial) conjugacy of the all the prior distributions, this adaptation is relatively straightforward. The updates for $(\mathbf{\Sigma}_\theta, \mu^{pop}, \mathbf{\Sigma}_\mu)$ are performed by sampling directly from their respective conditional posteriors, while the update steps for $(\boldsymbol{\theta}, \mu^{(1)}, \mu^{(2)})$ remain as before. The full

algorithm is given in Algorithm 3.1.

The partial pooling group structure is our favoured approach as it permits group-level comparisons (unlike complete pooling) while still pooling information across groups (unlike no pooling). We therefore performed the main analyses for the partial pooling approach only, though we also compared the no pooling and complete pooling approaches for the networks constructed via 10% absolute thresholding.

Proposals

For reasons of computational efficiency discussed in Section 2.5.2, we used proposal functions based on individual Bayesian ERGM fits of the networks in the two groups. For each network $\mathbf{y}^{(i,j)}$, we produce $K = 12000$ samples $\{\theta_k^{(i,j)}\}_{k=1}^K$ from their respective posterior distributions $\pi(\theta^{(i,j)}|\mathbf{y}^{(i,j)})$. In particular, we obtain estimates $\bar{\theta}^{(i,j)} = \frac{1}{K}\sum_{k=1}^K \theta_k^{(i,j)}$ of the mean and

$$\hat{\mathbf{\Sigma}}^{(i,j)} = \frac{1}{K-1}\sum_{k=1}^{K}(\theta_k^{(i,j)} - \bar{\theta}^{(i,j)})(\theta_k^{(i,j)} - \bar{\theta}^{(i,j)})^T$$

of the covariance for the posterior distribution from each network. We also obtain the group-level mean

$$\bar{\mu} = \frac{1}{n_1+n_2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\bar{\theta}^{(i,j)}.$$

For the individual-level proposals, we draw $\theta' \sim \mathcal{N}(0, \frac{\tau_\theta}{n_1+n_2}\sum_{j=1}^2\sum_{i=1}^{n_j}\hat{\mathbf{\Sigma}}^{(i,j)})$ where $\tau_\theta > 0$ is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4, as discussed in Section 2.5.2. For the group-level proposals, we calculate

$$\mathbf{\Sigma}_{\mu,prop} = \frac{1}{n_1+n_2-1}\sum_{j=1}^{2}\sum_{i=1}^{n_j}(\bar{\theta}^{(i,j)} - \bar{\mu})(\bar{\theta}^{(i,j)} - \bar{\mu})^T$$

and draw proposals $\mu' \sim \mathcal{N}(\mu, \tau_\mu\Sigma_{\mu,prop})$ where again $\tau_\mu > 0$ is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4.

Initialisation

To reduce the burn-in period, we initialised the algorithm at values based on the individual fits described above. Specifically, we set $\theta_0^{(i,j)} = \bar{\theta}^{(i,j)}$ and $\mu_0^{pop} = \bar{\mu}$. We then generated initial values for the group-level means $\boldsymbol{\mu}_0 = (\mu_0^{(1)}, \mu_0^{(2)})$ from the prior.

Algorithm 3.1 The ancillarity-sufficiency interweaving strategy (ASIS) algorithm for a two-group multilevel Bayesian ERGM

Require: number of MCMC iterations N, initial values $(\mu_0^{pop}, \boldsymbol{\mu}_0, \tilde{\boldsymbol{\theta}}_0)$

for $k = 1, \ldots, N$ **do**

- draw $\boldsymbol{\Sigma}_{\theta,k} \sim \pi(\cdot|\tilde{\boldsymbol{\theta}}_{k-1})$
- draw $\boldsymbol{\Sigma}_{\mu,k} \sim \pi(\cdot|\boldsymbol{\mu}_{k-1}, \mu_k^{pop})$
- draw $\mu_k^{pop} \sim \pi(\cdot|\boldsymbol{\Sigma}_{\mu,k}, \boldsymbol{\mu}_{k-1})$

for $j = 1, 2, \quad i = 1, \ldots, n_j$ **do**

- set $\theta_{k-1}^{(i,j)} = \mu_{k-1}^{(j)} + \tilde{\theta}_{k-1}^{(i,j)}$
- draw $\theta' \sim h_\theta(\cdot|\theta_{k-1}^{(i,j)})$
- draw $\mathbf{y}' \sim \pi(\cdot|\theta')$
- set $\theta_{k-0.5}^{(i,j)} = \theta'$ with probability
  $\min\left(1, AR(\theta', \theta_{k-1}^{(i,j)}; \mu_{k-1}^{(j)}, \boldsymbol{\Sigma}_{\theta,k}, \mathbf{y}^{(i,j)}, \mathbf{y}')\right)$ ▷ See Eq. (2.31)
- else, set $\theta_{k-0.5}^{(i,j)} = \theta_{k-1}^{(i,j)}$

end for

- draw $\mu_{k-0.5}^{(j)} \sim \pi(\cdot|\mu_k^{pop}, \boldsymbol{\theta}_{k-0.5}^{(j)}, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{\Sigma}_{\mu,k})$ j = 1,2

for $j = 1, 2, \quad i = 1, \ldots, n$ **do**

- set $\tilde{\theta}_k^{(i,j)} = \theta_{k-0.5}^{(i,j)} - \mu_{k-0.5}^{(j)}$

end for

for $j = 1, 2$ **do**

- draw $\mu' \sim h_\mu(\cdot|\mu_{k-0.5}^{(j)})$
- draw $\mathbf{y}^{(i)'} \sim \pi(\cdot|\mu' + \tilde{\theta}_k^{(i,j)})$ for $i = 1, \ldots, n_j$
- set $\mu_k^{(j)} = \mu'$ with probability
  $\min\left(1, AR(\mu', \mu_{k-0.5}^{(j)}; \tilde{\boldsymbol{\theta}}_k, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{y}^{(j)}, \boldsymbol{y}')\right)$ ▷ See Eq. (2.33)
- else, set $\mu_k^{(j)} = \mu_{k-0.5}^{(j)}$

end for

end for

Priors

We used informative values for the hyperparameters $(\mu_0, \mathbf{\Psi}_\theta, \mathbf{\Psi}_\mu)$, based on the individual Bayesian ERGM fits for all 652 individuals in the fMRI dataset. Writing $\{\theta_k^{(i)}\}_{k=1}^K$ for the $K = 12000$ posterior samples based on the network of the i^{th} individual (ignoring any group membership), we obtained estimates $\bar{\theta}^{(i)} = \frac{1}{K}\sum_{k=1}^K \theta_k^{(i)}$ of the posterior mean for each individual. Writing $\bar{\mu} = \frac{1}{652}\sum_{i=1}^{652} \bar{\theta}^{(i)}$ and $\hat{\mathbf{\Sigma}} = \frac{1}{651}\sum_{i=1}^{652}(\bar{\theta}^{(i)} - \bar{\mu})(\bar{\theta}^{(i)} - \bar{\mu})^T$, we then set $\mu_0 = \bar{\mu}$ and $\mathbf{\Psi}_\theta = \mathbf{\Psi}_\mu = \hat{\mathbf{\Sigma}}$. For the remaining hyperparameters, we set $\mathbf{\Lambda} = 5I$ and $\nu_\theta = \nu_\mu = p + 1$. Recall that p is the number of summary statistics, so $p = 2$ for the proportional thresholding case, and $p = 3$ for absolute thresholding.

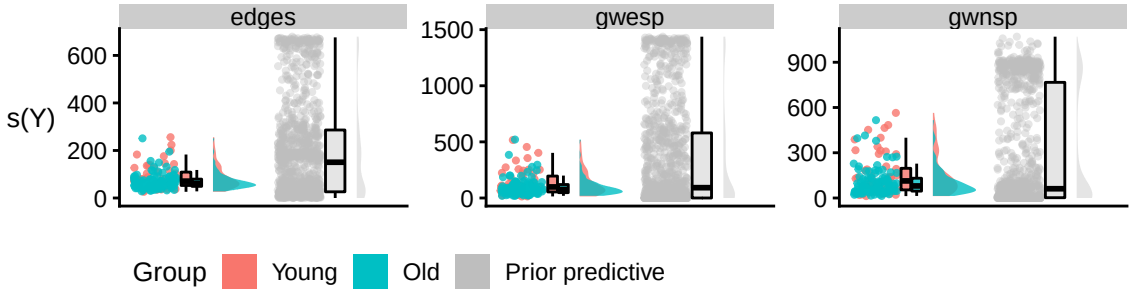
We assessed the above prior specification using prior predictive checks (Figures 3.4 and 3.5; see Section 2.4.2 for details). For both thresholding approaches and threshold values, the prior distribution for μ^{pop} generates networks with a wide range of summary statistic values that cover those in the observed dataset, indicating that the hyperparameters are appropriate. Similarly, the populations of networks generated from the prior distribution for $\mathbf{\Sigma}_\theta$ exhibit a range of spreads that appear compatible with the observed data.

The majority of the networks in the prior predictive populations (grey plots) have similar summary statistics to those of the observed data for the young group (red plots) and the old group (blue plots). Some of the populations, however, also contain networks with vastly different summary statistics, suggesting the prior is not overinformative.

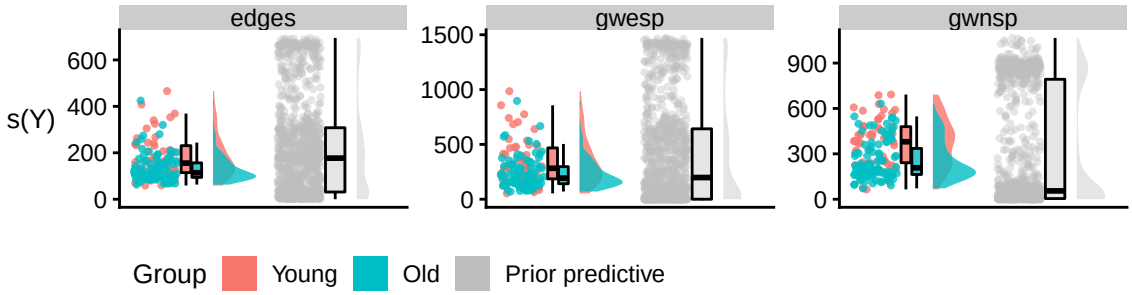
3.3.2 Detecting group differences

One of our objectives is to determine with a degree of certainty whether there are differences in the network structure between the two groups. There are a number of ways to approach this task. Our approach is to compare the posterior samples of the group-level mean parameters. Given samples from the joint posterior distribution of the parameters, we can construct samples of the difference between the group parameters and use these to determine whether there is a difference between the groups. We also considered alternative approaches such as the widely applicable information criterion (WAIC) [160] and Bayes factors, though we did not use these due to the considerable computational cost involved.

(a) 10% threshold - μ^{pop}



(b) 20% threshold - μ^{pop}



(c) 10% threshold - $\mathbf{\Sigma}_\theta$

(d) 20% threshold - $\mathbf{\Sigma}_\theta$

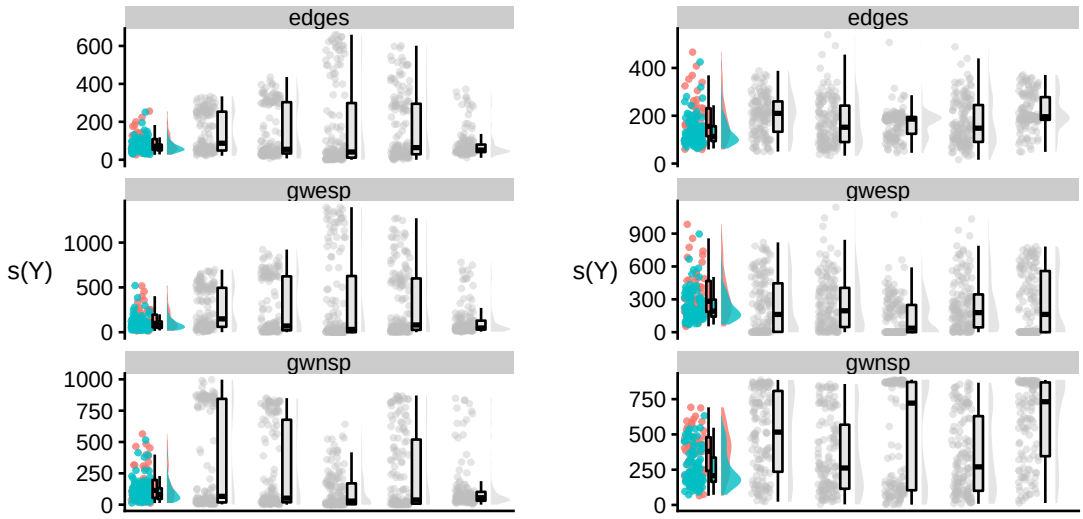
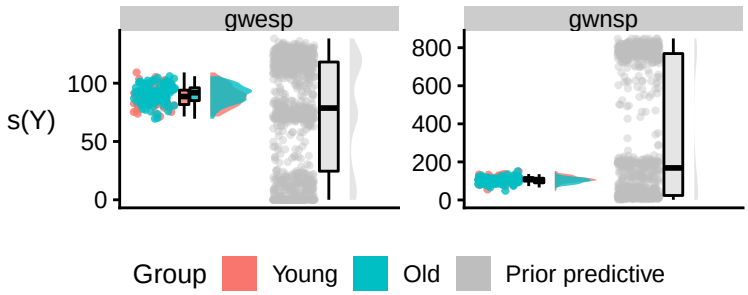
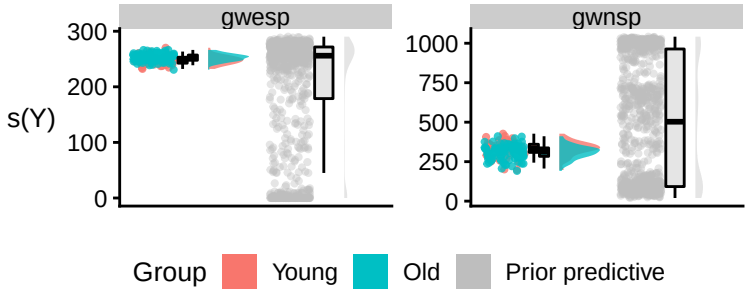


Figure 3.4: Prior predictive checks for absolute thresholding. The coloured plots correspond to the observed data. The grey plots correspond to networks generated under the respective prior predictive distributions. For μ^{pop}, the range of prior predictive values covers those of the observed data for both a) the 10% threshold, and b) the 20% threshold, indicating an appropriate prior distribution. Similarly, the populations of networks generated from the prior distribution for $\mathbf{\Sigma}_\theta$ exhibit a range of spreads that appear compatible with the observed data for both c) the 10% threshold, and d) the 20% threshold.

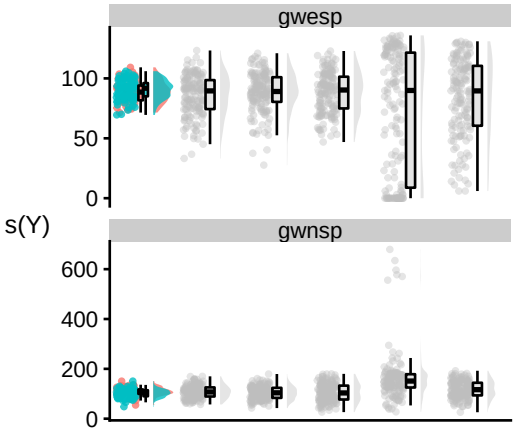
(a) 10% threshold - μ^{pop}



(b) 20% threshold - μ^{pop}



(c) 10% threshold - $\mathbf{\Sigma}_\theta$



(d) 20% threshold - $\mathbf{\Sigma}_\theta$

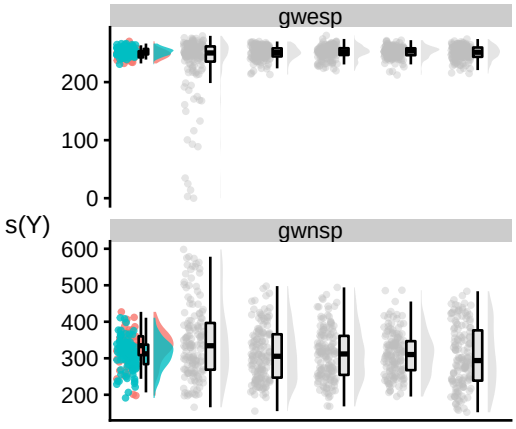


Figure 3.5: Prior predictive checks for proportional thresholding. The coloured plots correspond to the observed data. The grey plots correspond to networks generated under the respective prior predictive distributions. For μ^{pop}, the range of prior predictive values covers those of the observed data for both a) the 10% threshold, and b) the 20% threshold, indicating an appropriate prior distribution. Similarly, the populations of networks generated from the prior distribution for $\mathbf{\Sigma}_\theta$ exhibit a range of spreads that appear compatible with the observed data for both c) the 10% threshold, and d) the 20% threshold.

Posterior samples

One way of comparing the group connectivity structures is to consider the difference between the group-level mean parameters $\mu^{(1)} - \mu^{(2)}$. Note that our MCMC algorithm produces samples $(\mu_1^{(1)}, \mu_1^{(2)}), \ldots, (\mu_N^{(1)}, \mu_N^{(2)})$ from the joint posterior $\pi(\mu^{(1)}, \mu^{(2)}|\boldsymbol{y})$. We may then take $\mu_1^{(1)} - \mu_1^{(2)}, \ldots, \mu_N^{(1)}, -\mu_N^{(2)}$ as samples from the posterior of the difference between the parameters $\mu^{(1)} - \mu^{(2)}|\boldsymbol{y}$. This allows us to test statistically whether there is a difference in connectivity structure between the two groups. For example, if the zero-vector lies outside the 95% credible set of $\mu^{(1)} - \mu^{(2)}|\boldsymbol{y}$, then this constitutes strong evidence of a global difference in connectivity between groups; or a component-wise difference if any one element excludes zero.

Bayes factors

Bayes factors provide a quantification of the support of one model over another given the data. Given two proposed models m_1, m_2 the Bayes factor is defined as:

$$BF_{m_1,m_2} = \frac{\pi(\boldsymbol{y}; m_1)}{\pi(\boldsymbol{y}; m_2)} \tag{3.13}$$

where $\pi(\boldsymbol{y}; m_k)$ is the evidence under model m_k. To determine whether there is a two-group structure to the data, we can compare the complete pooling model (m_1; one group) to either the partial pooling model or the no pooling models (m_2; two groups). A large value of BF_{m_1,m_2} is evidence of a one-group structure.

In practice, however, calculating the model evidence is difficult. Chib's approach [29] is to write the evidence as:

$$\pi(\boldsymbol{y}; m_k) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}^*; m_k)\pi(\boldsymbol{\theta}^*; m_k)}{\pi(\boldsymbol{\theta}^*|\boldsymbol{y}; m_k)} \tag{3.14}$$

for any $\boldsymbol{\theta}^*$ in $\boldsymbol{\Theta} = \mathbb{R}^{p \times n}$. One can typically calculate $\pi(\boldsymbol{y}|\boldsymbol{\theta}^*; m_k)$ and $\pi(\boldsymbol{\theta}^*; m_k)$ analytically. Then, all that remains is to estimate the posterior $\pi(\boldsymbol{\theta}^*|\boldsymbol{y}; m_k)$ for some value $\boldsymbol{\theta}^*$ in order to obtain an estimate for the model evidence.

However, in our case, we cannot calculate the sampling distribution $\pi(\boldsymbol{y}|\boldsymbol{\theta}^*; m_k)$ due to the presence of the intractable normalising constant. Fortunately, we can bypass this issue when calculating the Bayes factor by noting that the sampling distribution is the same under the both models, i.e. $\pi(\boldsymbol{y}|\boldsymbol{\theta}^*; m_1) = \pi(\boldsymbol{y}|\boldsymbol{\theta}^*; m_2)$. As we have chosen conjugate hyperpriors, we can obtain analytical expressions for $\pi(\boldsymbol{\theta}^*; m_k)$ under both models.

Thus, to estimate BF_{m_1,m_2}, we need only get estimates for $\pi(\boldsymbol{\theta}^*|\boldsymbol{y}; m_2)/\pi(\boldsymbol{\theta}^*|\boldsymbol{y}; m_1)$.

To this end, we can use the 'Importance-weighted marginal Bayesian posterior density estimation' approach [27] to estimate the $\pi(\boldsymbol{\theta}^*|\boldsymbol{y}; m_k)$ separately. This uses the output from the MCMC algorithm in order to estimate the marginal posterior $\pi(\boldsymbol{\theta}^*|\boldsymbol{y}; m_k)$ Dropping the model subscript for notational convenience, we can estimate for the one-group model:

$$\hat{\pi}(\boldsymbol{\theta}^*|\boldsymbol{y}) = \frac{1}{N}\sum_{k=1}^{N} w(\boldsymbol{\theta}_k|\mu_k, \boldsymbol{\Sigma}_k)\frac{\pi(\boldsymbol{\theta}^*, \mu_k, \boldsymbol{\Sigma}_k|\boldsymbol{y})}{\pi(\boldsymbol{\theta}_k, \mu_k, \boldsymbol{\Sigma}_k|\boldsymbol{y})} \tag{3.15}$$

where the $\boldsymbol{\theta}_k, \mu_k, \boldsymbol{\Sigma}_k$ are posterior samples (i.e. from the MCMC algorithm), N is the number of MCMC iterations, and w is a weight function. Observe that we can write

$$\begin{aligned}\frac{\pi(\boldsymbol{\theta}^*, \mu_k, \boldsymbol{\Sigma}_k|\boldsymbol{y})}{\pi(\boldsymbol{\theta}_k, \mu_k, \boldsymbol{\Sigma}_k|\boldsymbol{y})} &= \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*|\mu_k, \boldsymbol{\Sigma}_k)}{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k)\pi(\boldsymbol{\theta}_k|\mu_k, \boldsymbol{\Sigma}_k)} \\ &= \exp\left\{\sum_i (\theta^{(i)*} - \theta_k^{(i)})^T s(\mathbf{y}^{(i)})\right\} \prod_i \left(\frac{Z(\theta_k^{(i)})}{Z(\theta^{(i)*})}\right) \frac{\pi(\boldsymbol{\theta}_k|\mu_k, \boldsymbol{\Sigma}_k)}{\pi(\boldsymbol{\theta}^*|\mu_k, \boldsymbol{\Sigma}_k)}.\end{aligned} \tag{3.16}$$

Unfortunately the $Z(\theta_k^{(i)})/Z(\theta^{(i)*})$ terms are intractable. However, noting that

$$\mathbb{E}_{\theta^{(i)*}}\left[\frac{\exp\{\theta_k^{(i)T} s(\mathbf{y})\}}{\exp\{\theta^{(i)*T} s(\mathbf{y})\}}\right] = \frac{Z(\theta_k^{(i)})}{Z(\theta^{(i)*})}, \tag{3.17}$$

we can generate ERGs $\mathbf{y}_j \sim \pi(\cdot|\theta^{(i)*})$ and replace the $Z(\theta_k^{(i)})/Z(\theta^{(i)*})$ by the Monte Carlo estimate $\frac{1}{J}\sum_{j=1}^{J} \exp\{(\theta_k^{(i)} - \theta^{(i)*})^T s(\mathbf{y}_j)\}$.

We find in practice, however, that this procedure yields highly unstable estimates of the posterior $\pi(\boldsymbol{\theta}^*|\boldsymbol{y})$ that are also highly sensitive to the choice of weight function w. For these reasons, as well as the computational cost of generating further ERGs, we do not pursue this approach further.

Widely applicable information criterion

Another alternative to model comparison is to use an information criterion. An information criterion is typically a measure of (expected) predictive accuracy. Since larger models will tend to have better predictive accuracy by virtue of their increased flexibility, some adjustment for the number of parameters is made to encourage model parsimony.

Here, we consider applying the widely applicable (or Watanabe-Aikake) information

criterion [160]. This is defined as:

$$WAIC = \sum_{i=1}^{n} \log \mathbb{E}[\pi(\mathbf{y}^{(i)}|\boldsymbol{\theta}; \boldsymbol{y})] - \sum_{i=1}^{n} \mathrm{Var}(\log \pi(\mathbf{y}^{(i)}|\boldsymbol{\theta}; \boldsymbol{y})) \tag{3.18}$$

where both the variance and the expectation are taken under the posterior distribution. Generally, one cannot evaluate either the expectation or the variance analytically so they are replaced by estimates based on the MCMC output of the fitting algorithm:

$$\widehat{WAIC} = \sum_{i=1}^{n} \log \left(\frac{1}{N} \sum_{k=1}^{N} \pi(\mathbf{y}^{(i)}|\theta_k^{(i)}) \right) - \sum_{i=1}^{n} \widehat{\mathrm{Var}}(\log \pi(\mathbf{y}^{(i)}|\theta^{(i)})) \tag{3.19}$$

where $\theta_k^{(i)}$ are MCMC samples and $\widehat{\mathrm{Var}}$ denotes the sample variance. The difficulty in using this approach is that we cannot compute the sampling density $\pi(\mathbf{y}^{(i)}|\theta_k^{(i)})$ due to the presence of the intractable normalising constant. While it may be possible to estimate the density at each step (e.g. using ideas discussed above), this would be highly impractical due to the computational cost of generating exponential random graphs.

3.4 Results

To investigate group differences in connectivity structure, we focused on the posterior samples of $\mu^{(1)}$ and $\mu^{(2)}$. Visual inspection of trace plots suggested good mixing of the chains. Recall that we considered two different network construction methods (absolute threshold and proportional threshold) with two threshold values each (10% and 20%). Not surprisingly, the findings vary depending on both the network construction approach and the threshold value.

Group differences

The posterior distributions for the group-level parameters reveal differences between the young group and the old group (Figure 3.6). For the 10% absolute threshold, the posterior samples for the GWNSP parameter tended to be larger for the young group than the old group, while the differences for the remaining two parameters are not as clear. For the 20% absolute threshold, posterior samples for both the GWNSP and the edges parameters were generally larger for the young group, though the GWESP parameter exhibited smaller values. For the 10% proportional threshold, the posterior samples for

GWNSP parameter again tended to be larger for the young group, while the samples for the GWESP parameter are very similar across groups. In contrast, for the 20% threshold, there appeared to be very little difference in the GWNSP parameter between the two groups, though the GWESP parameter tended to exhibit smaller values for the young group. Note that there were no posterior samples for the edges parameter because the number of edges was held fixed.

While there appeared to be clear separation between the groups for some of the ERGM parameters, the 95% credible intervals for each individual parameter of $\mu^{(1)} - \mu^{(2)}$ contained zero in every case except the GWESP parameter under the 20% absolute threshold (Figure 3.7). Note that these marginal credible intervals can be used to assess whether there is a group difference in a given summary statistic parameter. If the aim is to assess whether there is a group difference in overall connectivity structure, then a joint 95% credible region should be used instead.

To assess the effect of the different group structure specifications (no pooling, partial pooling, complete pooling), we generated posterior samples for each case under the 10% absolute threshold. The resulting posterior density estimates for the no pooling specification, in which the groups are considered completely independent, are highly similar to those for the partial pooling specification (Figure 3.8). As expected, the posterior samples under the complete pooling specification, in which the two groups are treated as one single group, sit between those of the young group and the old group under the no pooling specification.

Posterior predictive checks

To assess the goodness-of-fit of our model, we performed posterior predictive checks of the group-wise median of each of the summary statistics, as described in Section 2.4.4. Briefly, we simulated 100 populations of networks based on the posterior samples produced by the fitting algorithm. These populations of networks were generated by first drawing a population of individual-level parameters from $\mathcal{N}(\mu_k, \mathbf{\Sigma}_{\theta,k})$ where $(\mu_k, \mathbf{\Sigma}_{\theta,k})$ are samples from the posterior. We then generated a single exponential random graph for each individual-level parameter.

We took the (group) median of each of the summary statistics for each simulated population, yielding from the posterior predictive distribution of the median of the summary statistics. Under the absolute threshold at 10%, the posterior predictive distribution for the old group covered the observed data. On the other hand, the observed data for each

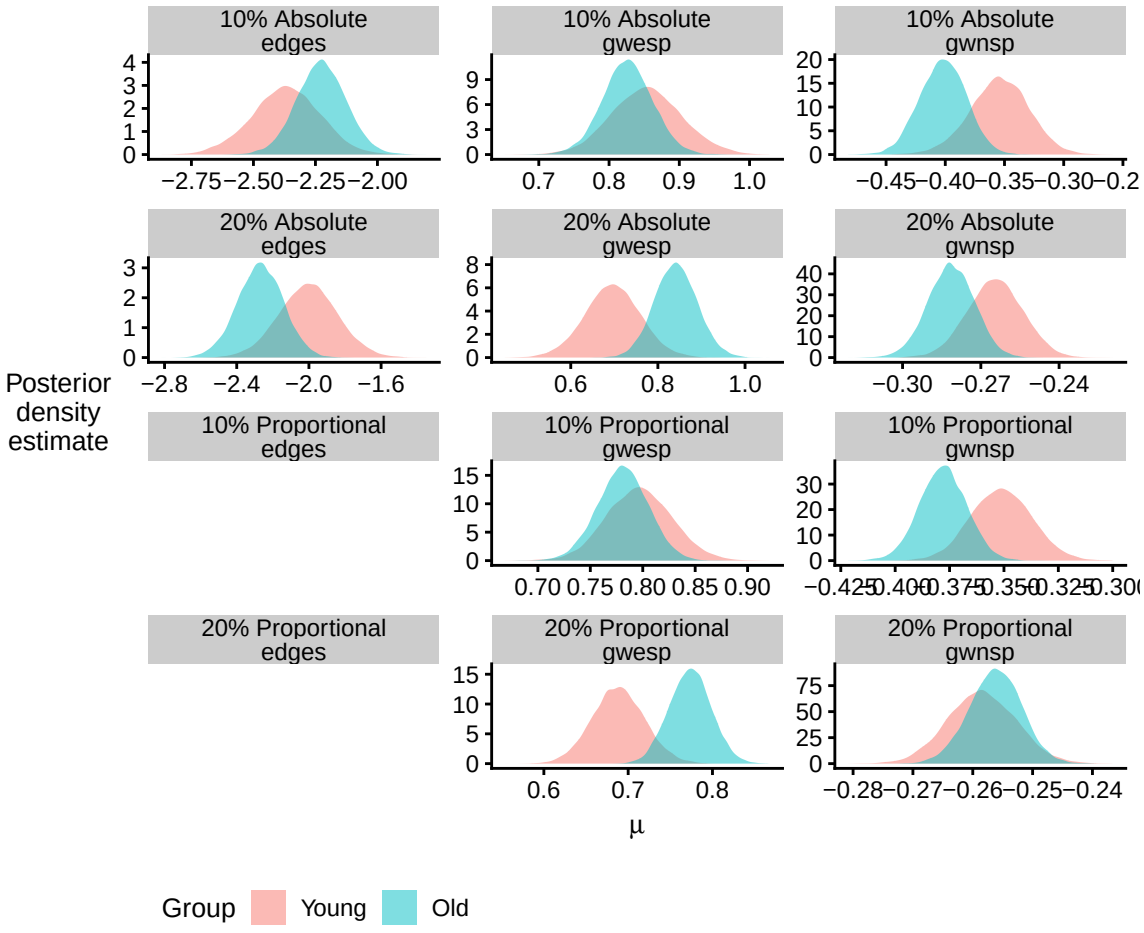


Figure 3.6: Posterior density estimates for the group-level mean parameters under the partial pooling group structure for all four thresholding approaches. Except for the 20% proportional threshold , the posterior placed more weight on higher values of the GWNSP for the young group (red plots) relative to the old group (blue plots). For both absolute and proportional thresholding at 20%, the posterior samples for the GWESP parameter tended to be smaller for the young group, while no such differences were seen at the 10% threshold. Group differences in the edges parameter were not as clear, with a tendency for larger values in the young group under the 20% absolute threshold.

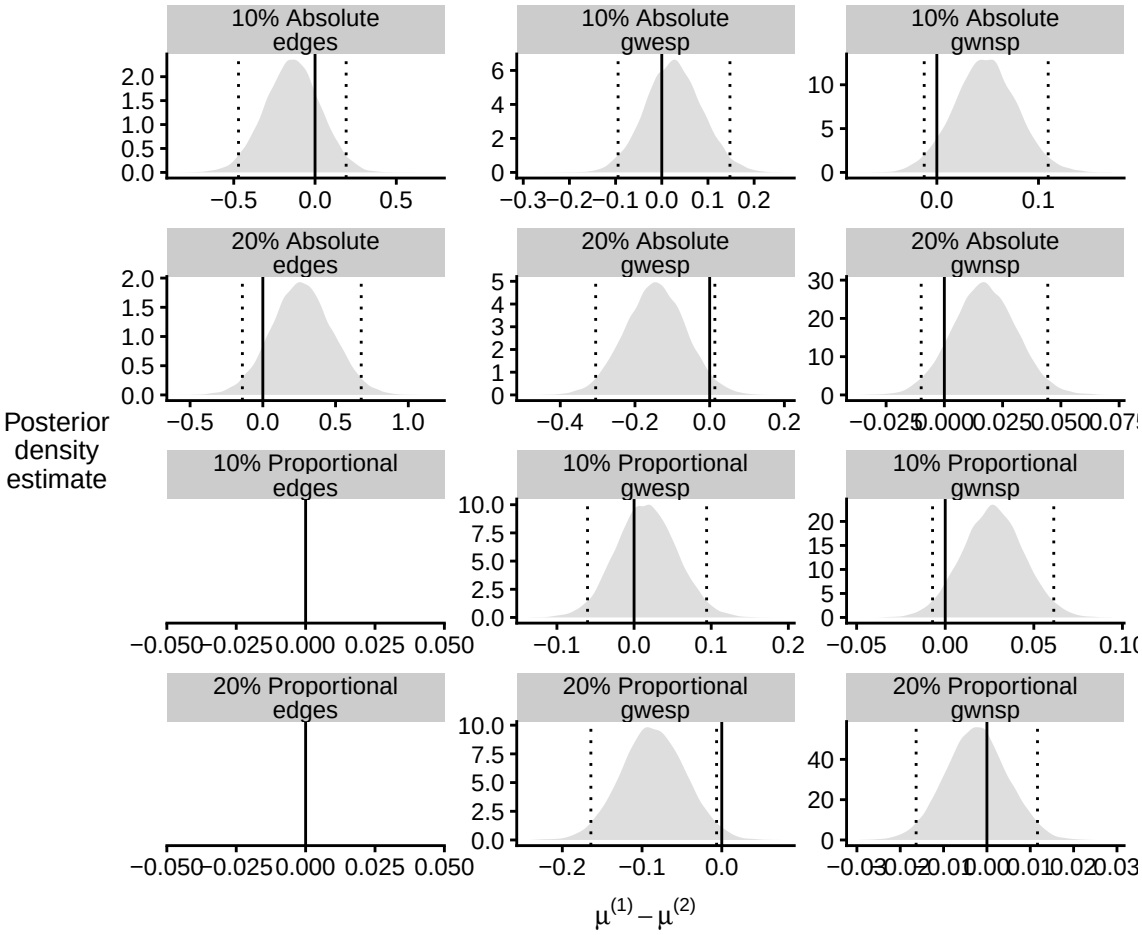


Figure 3.7: Posterior density estimates for the difference between the group-level mean parameters under the partial pooling group structure for all four thresholding approaches. The dotted lines indicate the 2.5% and 97.5% quantiles for each of the posterior samples, so that the gap between these lines corresponds to a 95% credible interval. The solid black line indicates zero. Except for the GWNSP parameter under the 20% absolute threshold, the 95% credible intervals contain zero for each of the parameters under each threshold procedure.

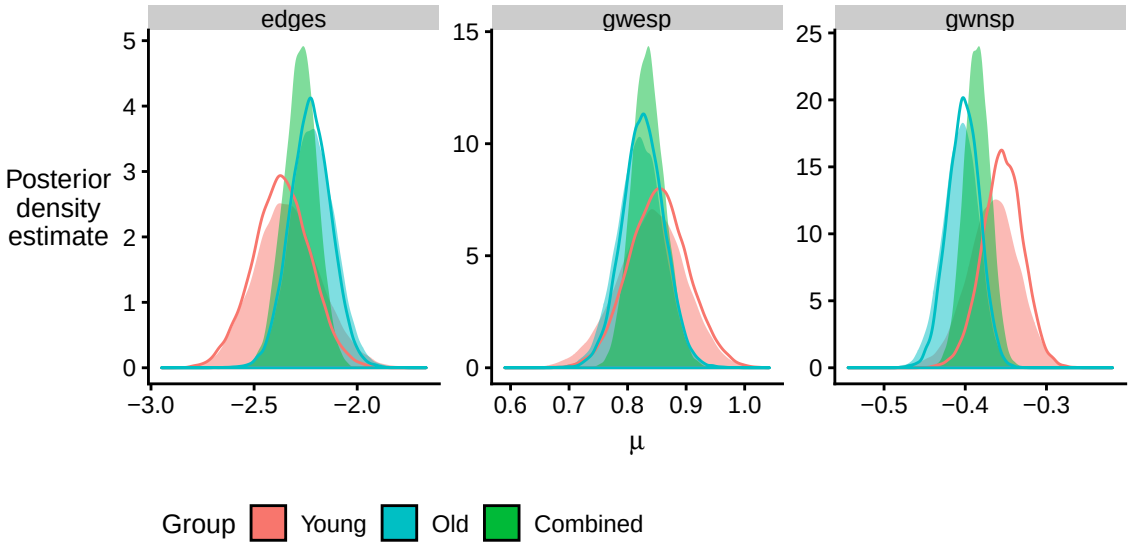


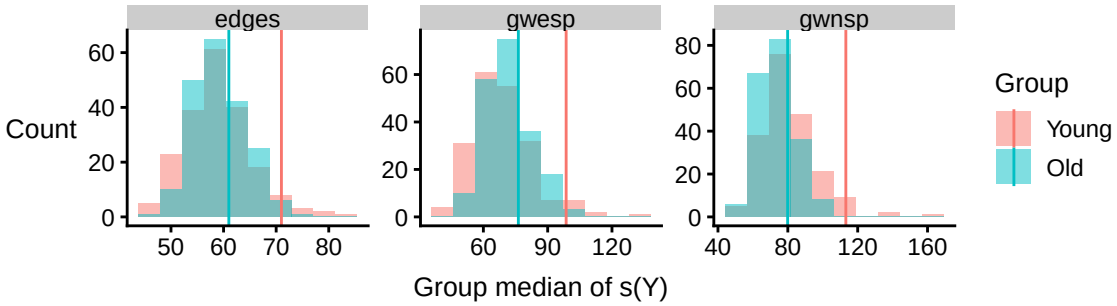
Figure 3.8: Posterior density estimates for the group-level mean parameters under the 10% absolute threshold with alternative group structure specification. For each of the three parameters, the corresponding posterior samples under the complete pooling structure (green plots) sit between those of the young group (red plots) and old group (blue plots) under the no pooling structure. The posterior density estimates under the partial pooling structure are indicated by the solid lines.

of the summary statistics of the young group sat in the tails of the posterior predictive distribution, indicating a poor fit (Figure 3.9a). For the absolute 20% threshold, only the GWNSP statistic of the young group sat in the tails of the posterior predictive distribution (Figure 3.9b). Under proportional thresholding, the posterior predictive check was more encouraging, with the posterior predictive distribution for both summary statistics covering that of the observed data (Figure 3.10).

We also compared the spread of the summary statistic values generated from the posterior predictive distribution with those of the observed data using raincloud plots (see Section 2.4.4) for details. Under the 10% absolute threshold, the posterior predictive populations overestimated the range of summary statistic values (Figure 3.11a). In particular, the simulated populations consistently included networks with larger values than those observed. Under the 20% threshold, the range of the summary statistics values was better captured by the posterior predictive distributions (Figure 3.11b). However, the posterior predictive populations did not consistently reflect the group-level differences observed in the data. For the proportional thresholding at both 10% and 20% the spread of summary statistic values from the posterior predictive populations closely matched that of the observed data, indicating a reasonable fit to the data (Figure 3.12).

While the posterior predictive distribution of the ERGM summary statistics appeared to match the observed data well under proportional thresholding, the same is not true in terms of other network metrics such as transitivity, associativity, and average path length (Figure 3.13). For example, observed median transitivity for both groups sat well in the tails of the corresponding posterior predictive distribution under the 20% proportional threshold.

(a) Absolute 10% threshold



(b) Absolute 20% threshold

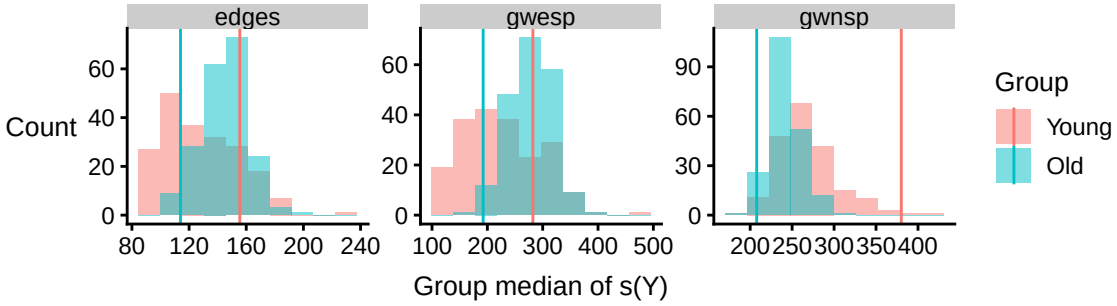
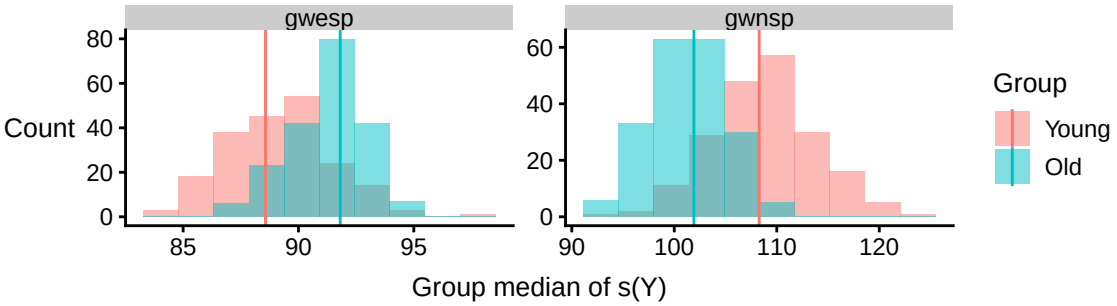


Figure 3.9: Posterior predictive check of groupwise summary statistic medians under absolute thresholding. The vertical lines correspond to the observed data. The histograms correspond to 100 populations of networks simulated from the posterior predictive distribution under group-level resampling. a) Under the 10% threshold, the posterior predictive distributions of the medians for the old group (blue plots) covered the observed data for each of the summary statistics. In contrast, the observed data for the young group (red plots) sat in the tails of the posterior predictive distribution, indicating a poor fit. b) Under the 20% threshold, the GWNSP statistic was not well covered by the posterior predictive distribution for both groups, nor the number of edges for the old group.

(a) Proportional 10% threshold



(b) Proportional 20% threshold

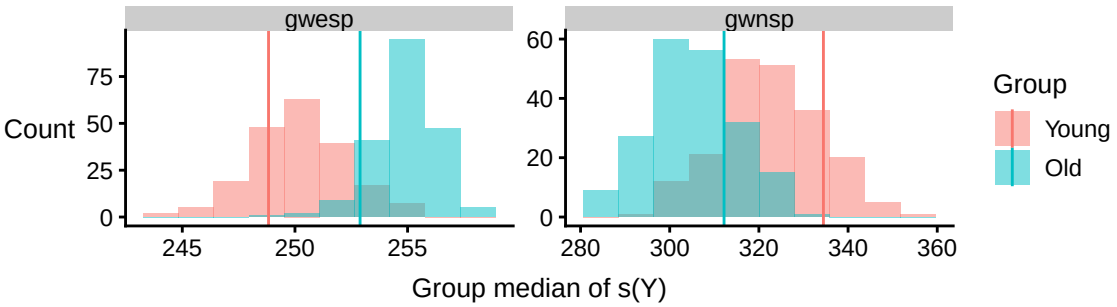
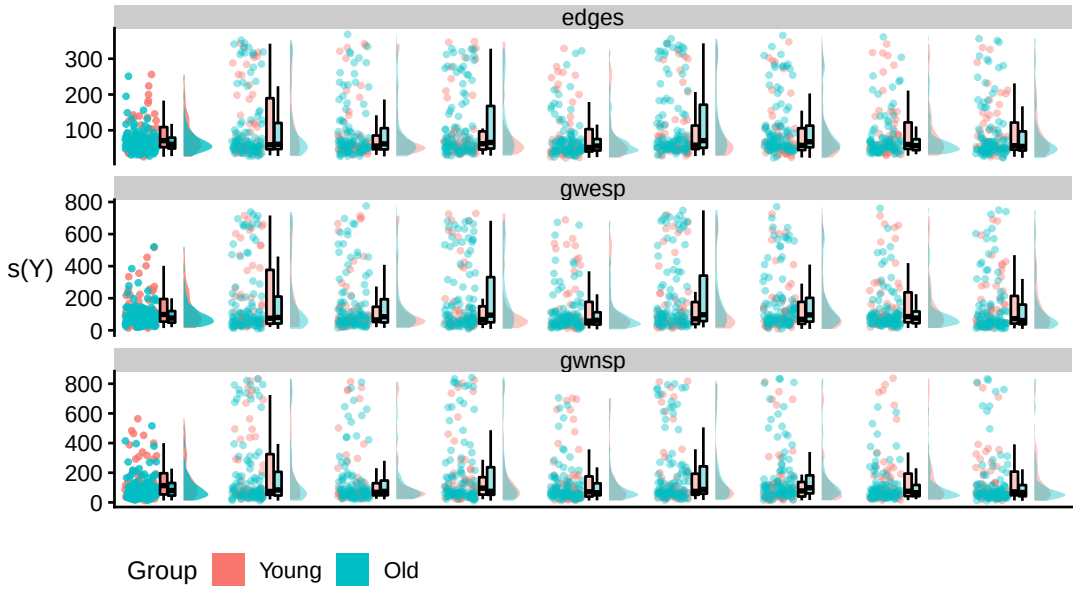


Figure 3.10: Posterior predictive check of groupwise summary statistic medians under proportional thresholding. The vertical lines correspond to the observed data. The histograms correspond to 100 populations of networks simulated from the posterior predictive distribution under group-level resampling. For the young group (red plots) and the old group (blue plots), the posterior predictive distribution of the summary statistic medians covered the observed data under both a) the 10% threshold and b) the 20% threshold.

(a) Absolute 10% threshold



(b) Absolute 20% threshold

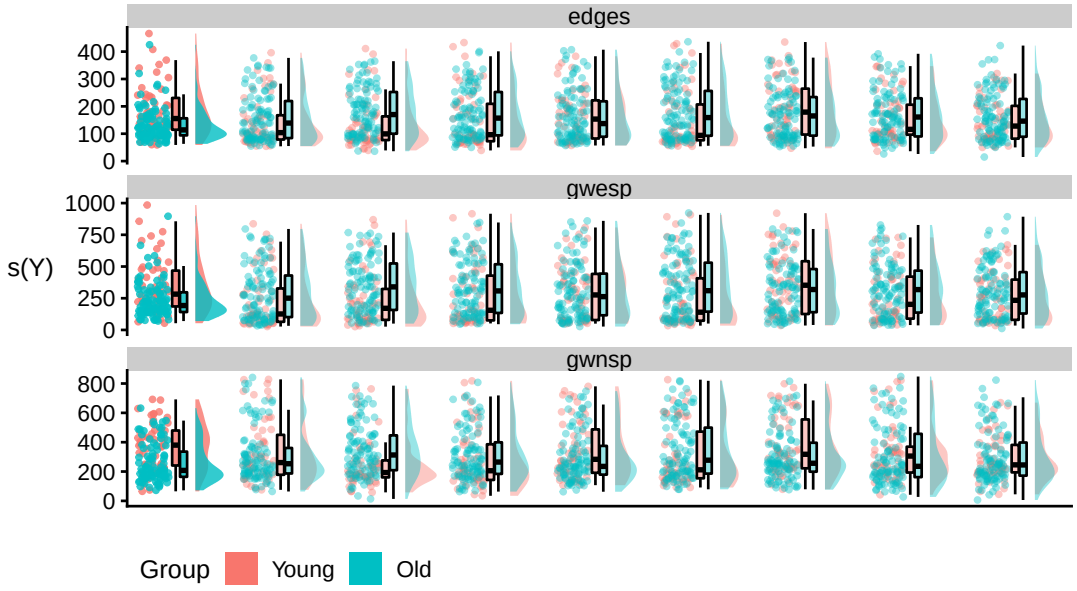
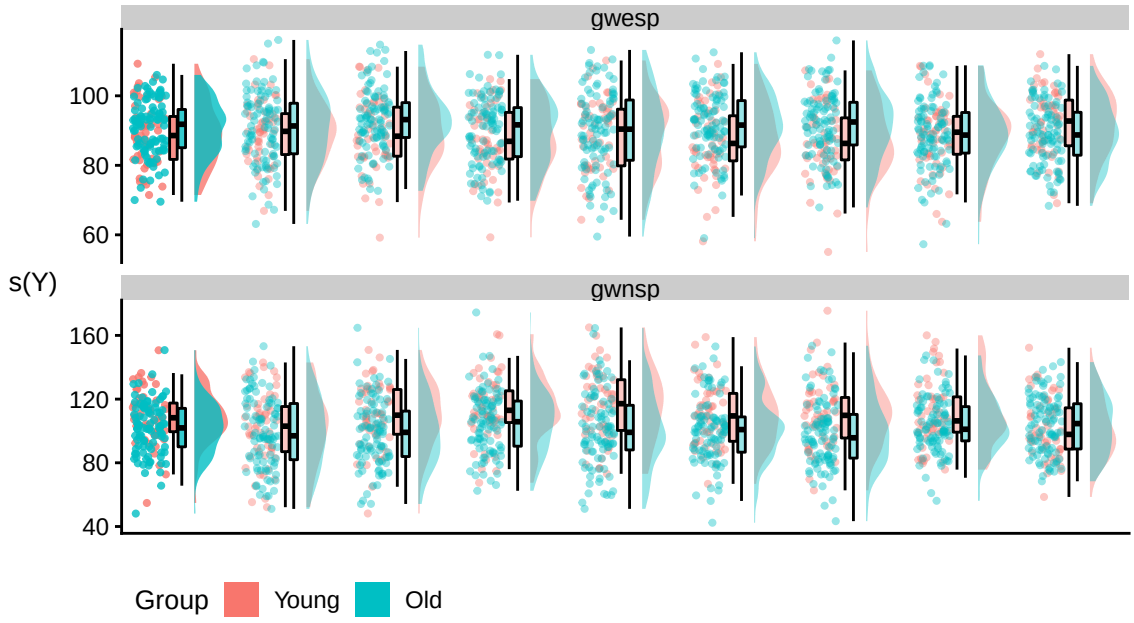


Figure 3.11: Posterior predictive check of spread of summary statistics for absolute thresholding. The leftmost plots correspond to the observed data, while the remaining plots correspond to populations of networks simulated from the posterior predictive distribution. a) Under the 10% threshold, the range of values of each of the three summary statistics was overestimated by the posterior predictive populations. b) Under the 20% threshold, the range of values was better captured by the posterior predictive populations. However, the posterior predictive populations did not consistently reflect the group-level differences in summary statistic values.

(a) Proportional 10% threshold



(b) Proportional 20% threshold

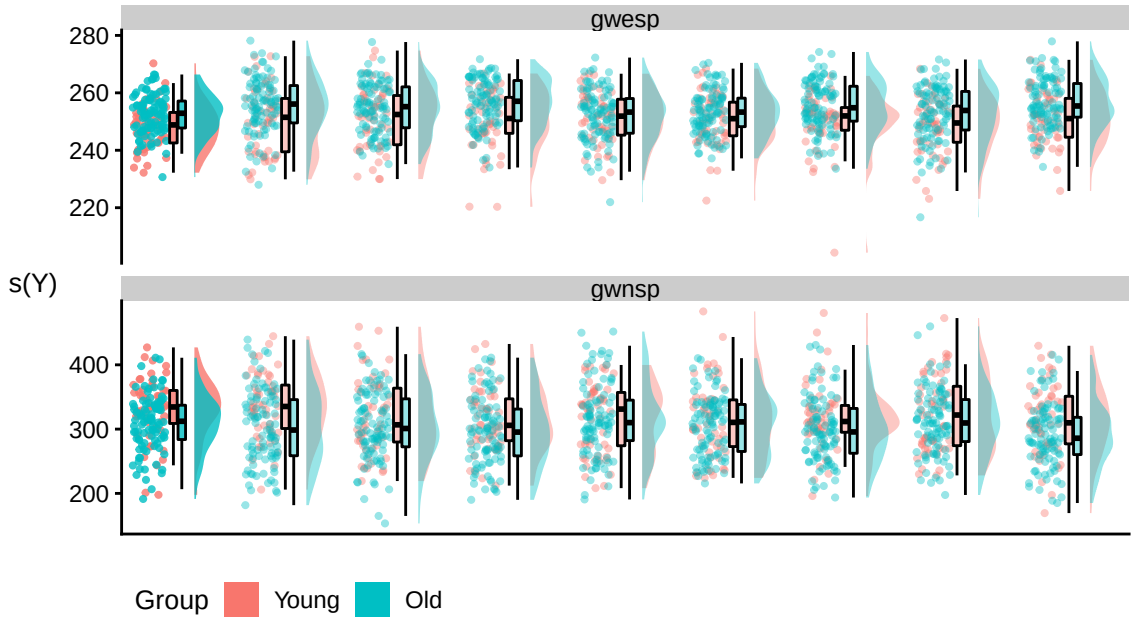
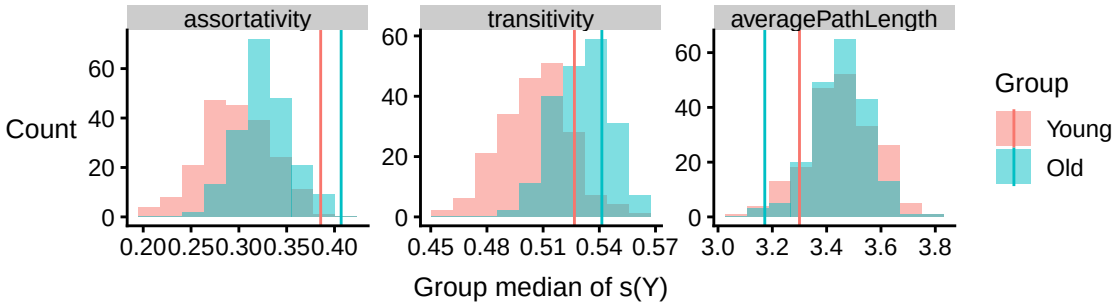


Figure 3.12: Posterior predictive check of spread of summary statistics for proportional thresholding. The leftmost plots correspond to the observed data, while the remaining plots correspond to populations of networks simulated from the posterior predictive distribution. The spread of summary statistic values of the posterior predictive populations matched the observed data reasonably well under both a) the 10% threshold, and b) the 20% threshold.

(a) Proportional 10% threshold



(b) Proportional 20% threshold

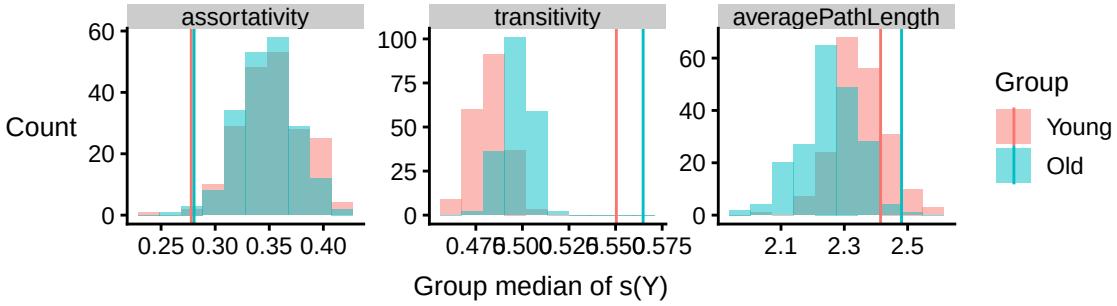


Figure 3.13: Posterior predictive check of groupwise medians of a selection of network metrics under proportional thresholding. The vertical lines correspond to the observed data. The histograms correspond to 100 populations of networks simulated from the posterior predictive distribution under group-level resampling. a) Under the 10% threshold, the posterior predictive distributions of median transitivity covered the observed data, though observed median assortativity sat in the tails of the distribution for both groups, indicating a poor fit. b) Under the 20% threshold, the observed median for all three metrics was not well covered by the posterior predictive distribution, with the exception of average path length for the young group.

3.5 Discussion

In this chapter, we have presented how the multilevel framework introduced in Chapter 2 can be applied to functional connectivity networks from two groups of individuals. The framework allows the pooling of information across individuals within the same group and, depending on the parametrisation, across groups. By analysing the posterior samples produced by the MCMC algorithm, we can determine whether there are statistically significant differences in the connectivity structure between the two groups. Here, we discuss some of the limitations of our analysis and identify some potential avenues for future work.

3.5.1 Limitations

Selection of summary statistics

There are a number of aspects to our approach that warrant further investigation. Firstly, from a modelling perspective, the appropriate selection of summary statistics is crucial. This is true of fitting ERGMs to single networks, let alone populations of networks. Moreover, the 'correct' choice likely depends on the network construction method. We used the summary statistics suggested in Simpson et al. [132], which were selected via a graphical goodness of fit approach.

The posterior predictive checks we performed indicated poor fit to the data. Under absolute thresholding, the group medians of the ERGM summary statistics were not well captured by the posterior predictive distributions. Under proportional thresholding, while the fit was reasonably good in terms of the ERGM statistics but not in terms of other network metrics explicitly included in the model. This suggests that these summary statistics were not flexible enough to adequately describe the networks constructed from the fMRI data. As a result, any group differences, or lack thereof, found in the parameters should be treated with caution. Determining which summary statistics should be included to yield a better fit is a challenging task, however, and needs to be investigated further.

Choice of prior distributions

Secondly, our current choice of hyperprior for the individual-level parameters is likely to be oversimplistic. Given the complex geometry of exponential random graph models [115], it is unlikely that the distribution of individual-level parameters is well described

by a multivariate normal. This is evidenced by the prior predictive plots: a large proportion of the networks simulated from the prior predictive distribution have vastly different values of summary statistics to those observed. A better understanding of the geometry of ERGMs is essential to construct better priors that constrain the individual-level parameters to appropriate regions of the parameter space.

Interpretation of parameter estimates

Thirdly, given the interdependencies inherent in the summary statistics, the interpretation of group differences in the parameters requires some care. In particular, it is important to consider the group differences across all the summary statistics simultaneously. For example, under the 20% absolute threshold, although the posterior samples for the edges parameter were generally lower for the old group, the median number of edges from the corresponding posterior predictive populations were generally higher.

This apparent peculiarity was likely due to the correlation of the summary statistics. For example, a higher value of GWESP generally entails a higher number of edges, resulting in a positive correlation in the two summary statistics. This leads to negative correlations between the corresponding parameters: a higher parameter for GWESP would generally correspond to networks with a higher value of GWESP and so a (much) higher number of edges. The edges parameter has to be smaller to compensate for this to maintain a similar level of sparsity in the network. Given these difficulties in interpretation, further work is needed to characterise a (posterior) distribution of ERGM parameters in order to quantify and visualise the variability of the networks on the scale of the summary statistics.

Computational cost

As mentioned previously, the computational cost of our MCMC algorithm is considerable. Using a 32-core computing cluster (Intel Xeon E7-8860 v3, 2.2 GHz), the algorithm took between 8 and 24 hours to produce the joint posterior samples for each of the datasets and models considered in this chapter. (This does not include the time taken to tune the proposal functions.) The main computational bottleneck lies in simulating the exponential random graphs at each MCMC iteration. While the computational cost should increase roughly linearly in the number of networks, Krivitsky & Handcock [86] provide empirical evidence suggesting the cost may grow on the order of $p(|V| + |E|)\log(|E|)$ where p is the number of summary statistics, $|V|$ is the number of nodes, and $|E|$ is the number of

edges. This limits the size of the networks that our framework can handle to relatively coarse parcellations of the brain (c. 100 regions).

Impact of thresholding procedure

Our results are dependent on the method used to construct the networks, reinforcing the importance of thresholding choice in connectivity analyses. At 10% the group differences were similar for absolute thresholding and for proportional thresholding, with marginal differences in the GWESP parameter, and tendency for higher values of the GWNSP parameter for the young group. In contrast, at 20% the group difference in the GWNSP parameter under absolute thresholding was not reflected under proportional thresholding.

Further, group differences were not consistent across threshold values. Under absolute thresholding at 20%, the posterior samples for the edges parameter were generally higher for the young group, while the posterior samples for the GWESP parameter were generally higher. At 10%, these differences were reversed, and not as clear-cut. Similarly, under proportional thresholding, the group difference in the GWNSP parameter present at 10% was not present at 20%, and vice versa for the group difference in the GWESP parameter.

3.5.2 Future work

Model selection

One of the key challenges in applying our framework to real data is the choice of which network summary statistics to include in the model. A fully Bayesian model selection method based on reversible-jump MCMC has been developed for exponential random graph models on single networks [23]. A similar approach could be developed for our framework, though the computational cost is likely to be prohibitive. A more pragmatic approach would be to develop a graphical goodness-of-fit method by comparing the posterior predictive distributions under different models.

Moreover, the choice of summary statistics should reflect our belief of how the global network structure may be summarised and should be driven by the context of the network. This is more difficult for brain networks than, for example, social networks, because we currently have a relatively limited understanding of the biological processes driving brain connectivity. Since most existing summary statistics were developed in the context

of social networks, more research is needed to develop and assess new summary statistics that better reflect the brain's connectivity structure.

Weighted networks

One way to bypass the choice of threshold would be to use weighted exponential random graph models [84, 38]. These are an extension of the binary ERGM that can be applied to weighted networks, thus avoiding the issue of thresholding. Indeed, one version of a weighted ERGM, the generalised exponential random graph model (GERGM) [38] was recently applied to a 20-node functional connectivity network [140]. This approach has the additional advantage of modelling the mean connectivity directly and thus would avoid any confounding due to differences in mean connectivity. However, the GERGM is at present extremely computational intensive, rendering it infeasible for a population of networks.

An alternative approach would be to explicitly account for noise in the model. Recently, a method [109] was proposed to reconstruct networks in the presence of noisy data by specifying the generative model for the network jointly with the noise measurement process. While this approach focused on the stochastic block model, the idea should still be applicable to ERGMs. Since higher levels of noise generally result in lower mean connectivity, accounting for the noise should lead to more reliable estimates of the overall connectivity structure.

Different group structures

The flexibility of the multilevel framework could be exploited to model more complex group structures such as multiple groups or factorial designs. Similarly, given multiple network measurements per individual, it is straightforward to extend the framework by adding another layer to the model. One could also incorporate individual-level covariate information into the model. In the context of healthy ageing, this would avoid splitting individuals into (somewhat arbitrary) groups and provide a mechanism to measure how connectivity structure varies throughout the lifespan.

Chapter 4

Assessing dynamic functional connectivity in heterogeneous samples*

Summary

Recent evidence indicates that static connectivity analyses may represent an average across informative fluctuations in functional connectivity. Several methods have been developed to measure dynamic functional connectivity (dFC) in fMRI data. These methods are often based on a sliding-window analysis, which aims to capture how the brain's functional organization varies over the course of a scan. The aim of many studies is to compare dFC across groups, such as younger versus older people. However, spurious group differences in measured dFC may be caused by other sources of heterogeneity between people. For example, the shape of the haemodynamic response function (HRF) and levels of measurement noise have been found to vary with age. We use a generic simulation framework for fMRI data to investigate the effect of such heterogeneity on estimates of dFC. Our findings show that, despite no differences in true dFC, individual differences in measured dFC can result from other (non-dynamic) features of the data, such as differences in neural autocorrelation, HRF shape, connectivity strength and measurement noise. We also find that common dFC methods can detect spurious group differences in dynamic connectivity due to inappropriate setting of their hyperparameters. fMRI studies therefore need to consider alternative sources of heterogeneity across individuals before concluding differences in dFC.

*__DISCLAIMER__: This work has been published in a journal article [89] from which the figures and large parts of the text of this Chapter are taken directly, with permission.

4.1 Background

Static functional connectivity is commonly measured using the Pearson correlation between the fMRI time series in different brain regions [19]. Typically, one would calculate the correlation between two time series over the course of the whole fMRI scan. However, this approach may represent an average across informative fluctuations in functional connectivity. Indeed, recent evidence suggests that even in task-free, resting-states these functional connections change over the course of a scan [7, 26, 80]. Moreover, measures of this dynamic functional connectivity (dFC) have been used in an attempt to identify biomarkers for schizophrenia [122]) and Alzheimer's disease [78].

The most common way to measure dFC is to apply a sliding-window analysis (see [77] for a review of dFC), which works by computing connectivity within successive windows of the fMRI time series. Suppose we have time series $(X_t)_{t=1}^T$ and $(Y_t)_{t=1}^T$ for two brain regions, where T is the length of the fMRI scan, and let w be the width of the sliding window. We will assume the window slides in time steps of $\tau = 1$. The functional connectivity for the j^{th} window $\rho^j_{X,Y}$ is then given by the Pearson correlation of the two time series truncated to this window:

$$\rho^j_{X,Y} = \frac{\sum_{t=j}^{j+w-1}(X_t - \bar{X}_j)(Y_t - \bar{Y}_j)}{\sqrt{\sum_{t=j}^{j+w-1}(X_t - \bar{X}_j)^2 \sum_{t=j}^{j+w-1}(Y_t - \bar{Y}_j)^2}} \tag{4.1}$$

where $\bar{X}_j = \frac{1}{n}\sum_{t=j}^{j+w-1} X_t$ and $\bar{y}_j = \frac{1}{n}\sum_{t=j}^{j+w-1} Y_t$ are the sample means of X_t and Y_t in the J^{th} window. This approach yields an estimate of connectivity between the two regions for each successive window, thus providing a measure for how this varies over the course of a scan. One can perform this analysis across all pairs of brain regions in the data to see how the whole-brain connectivity structure varies with time.

Two important parameters in the above description are the window width w and sliding-step size τ. The effect of these parameters on the estimates of dFC have previously been analysed in detail [128]. In brief, while smaller window sizes may capture brief changes in connectivity, they are more susceptible to spurious fluctuations. Sliding-step sizes of $\tau > 1$ may reduce the sensitivity of the sliding-window approach to any changes in connectivity, particularly if the time between changes is similar to τ.

One common adaptation to (4.1) is to use a tapered window (e.g. [7], [68]). By putting a larger weight on observations in the centre of the window , this reduces the effect of any artifacts at the edge of the window. For example, the tapered cosine, or *Tukey*, window

uses weights

$$\omega_t = \begin{cases} \frac{1}{2}\left[1+\cos\left(\pi\left(\frac{2t}{\alpha w}-1\right)\right)\right] & 1 \leq t < \frac{\alpha w}{2} \\ 1 & \frac{\alpha w}{2} \leq t \leq w(1-\frac{\alpha}{2}) \\ \frac{1}{2}\left[1+\cos\left(\pi\left(\frac{2t}{\alpha w}-\frac{2}{\alpha}+1\right)\right)\right] & w(1-\frac{\alpha}{2}) < t \leq w \end{cases} \tag{4.2}$$

where w is the window length, and α is a parameter controlling the length of the taper. Using a tapered window, the functional connectivity between two regions for the window is then given by the Pearson correlation of the truncated time series weighted by the ω_t.

By applying the above sliding-window approach, we can calculate a connectivity measure between all pairs of brain regions for windows through course of the scan. A variety of methods have been proposed to analyse the changes in connectivity across windows. A simple approach characterises dFC as the standard deviation (SD) of the correlation values across time windows [42]. For a given pair of regions X and Y, the standard deviation $\sigma_{corr(X,Y)}$ of the functional connectivity is given by:

$$\sigma_{corr(X,Y)} = \sqrt{\frac{\sum_{j=1}^{n-w+1} \rho_{X,Y}^{j} - \bar{\rho}_{X,Y}}{n-w}} \tag{4.3}$$

where $\bar{\rho}_{X,Y} = \frac{1}{n-w+1}\rho_{X,Y}^{j}$ is the mean connectivity across all the windows. A large value of $\sigma_{corr(X,Y)}$ indicates that the connectivity between X and Y is more variable.

An alternative, whole-brain approach is to pool data across individuals and use k-means clustering [7]. By characterising the connectivity structure of the i^{th} subject in the j^{th} window as a matrix $R^{ij} \in \mathbb{R}^{N\times N}$, where N is the number of brain regions, one can then apply k-means clustering to identify recurring connectivity patterns, or "FC states", across the R^{ij}. Another class of methods applies network theory on an individual level, with distinct brain regions corresponding to nodes, and functional connections corresponding to edges between nodes [22]. The multilayer modularity approach [15, 16] characterises the dynamic connectivity structure for an individual as a single multilayer network and then uses a community detection algorithm to identify modules of connected regions at different points during the scan.

One of the notable challenges for both static and dynamic FC methods is that of individual heterogeneity. While the aim of group studies is to detect heterogeneity in FC, other sources of heterogeneity may have an impact on estimates of FC. Heterogeneity can arise in a variety of ways. For example, the autocorrelation of fMRI time series

within brain regions differs between healthy brains and those with schizophrenia [10], and autocorrelation is known to affect estimation of cross-correlation [11]. Although it is unclear whether this change in autocorrelation is due to neural or vascular factors, work with dynamic causal modelling has suggested that neural autocorrelation within some networks can vary between young and older participants [149]. Another example of heterogeneity is differences in the haemodynamic response function (HRF). The shape of the HRF, which can be modelled as a finite impulse response kernel, has been found to vary between healthy patients and patients with schizophrenia [69] and also between age groups [73, 3, 39]. Even non-neural physiological noise levels might differ across groups, owing for example to greater within-scan head movement in old relative to young subjects [53].

The effect of heterogeneity on estimates of static FC has previously been studied in the context of healthy ageing [54]. Building on this work, here we aim to investigate how unaccounted heterogeneity impacts estimates of *dynamic* FC. Based on a previous study[7], we designed a simulation framework to generate data from a dynamic connectivity structure based on FC states. We characterise FC states as time periods in which brain regions can be grouped into specific sets, or "modules". In a given FC state, regions are considered connected if, and only if, they are in the same module. In this framework, changes in connectivity structure then correspond to FC state transitions. Data were generated to investigate the effect of individual differences in neural autocorrelation, HRF shape, connectivity strength and measurement noise on estimated dFC. We specifically used the case of ageing to illustrate how plausible age-related sources of heterogeneity could impact dFC estimates. Furthermore, we varied the number of FC states and frequency of FC state transitions, in order to explore the effect of hyperparameter selection on the results of popular dFC methods such as the k-means method [7] and the multilayer modularity approach [15]. Our analysis follows a typical sliding-window dFC pipeline based on a number of recent studies (e.g. [7, 42, 122]). Our findings show that group-level differences in neural autocorrelation, HRF shape, connectivity strength, measurement noise, number of FC states and frequencies of FC state transitions can lead to systematic differences in observed dFC between simulated fMRI time-series data.

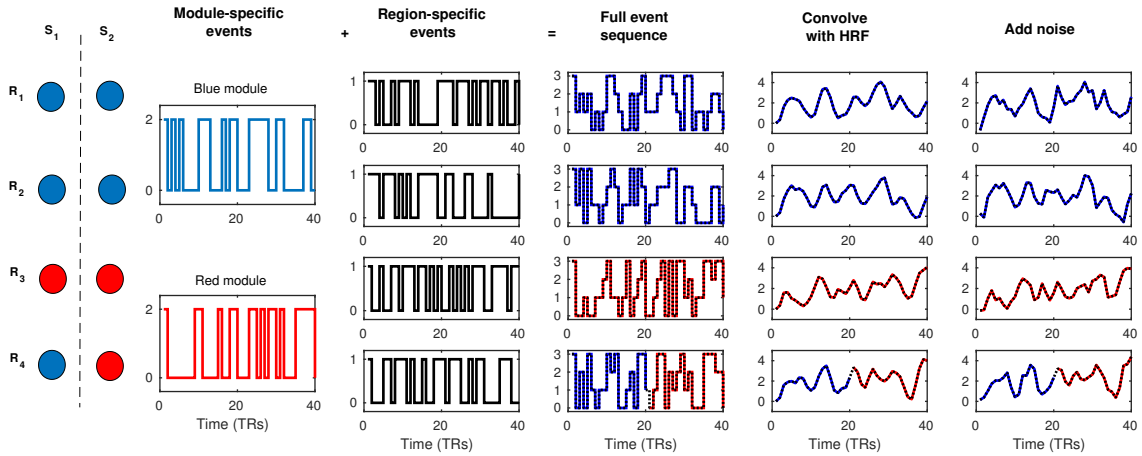


Figure 4.1: The generic simulation framework used to generate fMRI data for a brain consisting of four regions. State i is denoted by S_i and region j is denote by R_j. Two brain regions are connected at any given time if they are in the same module, which are distinguished here by colours. These panels show data for 40 TRs: the first 20 TRs are spent in S_1 while the last 20 TRs are spent in S_2.

4.2 Methods: simulation framework

To demonstrate some of the issues associated with assessing dFC in heterogeneous samples, we developed a simulation framework based on a previous study [7]. For each type of heterogeneity, we report the effects of changing only one parameter at a time, in order to isolate its relative impact on the analysis. Thus, each source of heterogeneity corresponded to a change in a single step of the data generation process. We also consider interactions between parameters to show how combinations of different sources of heterogeneity may impact measures of dFC. To illustrate the effects of neural autocorrelation, HRF, connectivity strength and measurement noise, it was sufficient to simulate data from a model with only four regions of interest (ROIs) - Base Simulation 1. In order to analyse the impact of changes in both number of FC states and frequency of FC state transitions, we increased the number of ROIs to 32 to allow for a greater variety of states - Base Simulation 2. We now describe the simulation framework in detail, illustrated schematically in Figure 4.1, before outlining both Base Simulations and the specific variations for each source of heterogeneity.

We characterised FC states as time periods in which regions are partitioned into "modules". We assumed that there is a finite number N_{states} of attainable FC states. For

convenience, we denote state i as S_i and region j as R_j. An FC state dictates the module membership for each region. We denote the module membership of region R_j in state S_i as $M_{i,j}$. To simulate fMRI data for an individual, we first generated a FC state sequence $H_t, \quad t = 1, \ldots, T$ where T is the number of samples in the scan, and H_t takes values in the set of states $S_1, \ldots, S_{N_{states}}$. This state sequence determines the module memberships of each region at each time point. The process of generating the state sequence is outlined in detail in the Base Simulations below.

We then used a sampling rate of TR = 2s and generated binary neural event sequences of length $T = 360$ TRs for each module and each region. The module-specific event sequences drive the connectivity structure: a module-specific event is one which occurs for all regions within that module in the current FC state. In contrast, region-specific events are those which occur for single regions only, independent of other regions, and thus correspond to neural noise. We denote the event sequences as time series $E_t^{(m)}$, $E_t^{(j)}$ taking values in $\{0, 1\}$ for module m and region R_j respectively.

More precisely, a module-specific event occurred in a module at an individual time point with probability $P_{mod} = 0.5$, independent of all other modules and time points. If a module has an event at time t, all regions within that module at time t have an event. For each region, we then superimposed a region-specific neural event sequence. A region-specific event occurred in a region at an individual time point with probability $P_{reg} = 0.5$, independent of all other regions and time points (except in Section 3.1, where we explored the effect of autocorrelated region-specific events). We fixed the amplitude of region-specific events to be $a_{reg} = 1$, and set the amplitude of the module-specific events to be $a_{mod} = 2$. The full event sequence for R_j can be written as:

$$Z_t^{(j)} = a_{reg} E_t^{(j)} + a_{mod} E_t^{(M_{H_t,j})} \tag{4.4}$$

where $M_{H_t,j}$ is the module membership of R_j in state H_t.

The full event sequences were then convolved with a haemodynamic response function (HRF) using the SPM12 software [47] (http://www.fil.ion.ucl.ac.uk/spm) to produce fMRI-like time series. We used the default HRF in SPM12, namely, the difference of two

Parameter	**Description**	**Default** (seconds)
p_1	Delay of response (relative to onset)	6
p_2	Delay of undershoot (relative to onset)	16
p_3	Dispersion of response	1
p_4	Dispersion of undershoot	1
p_5	Ratio of response to undershoot	6
p_6	Onset	0
p_7	Length of kernel	32

Table 4.1: The parameters of the default haemodynamic response function in SPM12 [47].

gamma functions. This can be written

$$h(t) = \begin{cases} \dfrac{(t-p_6)^{\frac{p_1}{p_3}-1} e^{-\frac{t-p_6}{p_3}}}{\Gamma(\frac{p1}{p3}) p_3^{\frac{p_1}{p_3}}} - \dfrac{1}{p_5} \dfrac{(t-p_6)^{\frac{p_2}{p_4}-1} e^{-\frac{t-p_6}{p_4}}}{\Gamma(\frac{p2}{p4}) p_4^{\frac{p_2}{p_4}}}, & \text{if } p_6 \leq t \leq p_7. \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

where $p_1, \ldots, p_7$ are a set of parameters governing the shape of the HRF (see Table 4.1). The time series were rescaled to have a SD of 1. White noise with SD $\sigma_{noise} = 0.2$ was then added. Finally, a high-pass Butterworth filter removing frequencies below 0.033Hz was applied. This is based on the rule of thumb which recommends removing frequency components below $1/w$, where w is the window length in the sliding-window analysis [90].

4.2.1 Base Simulation 1 (4 ROIs)

In this setting, which corresponds to the framework illustrated in Figure 4.1, we restricted dynamics to two FC states, S_1 and S_2. S_1 corresponded to the partition {1,1,2,1}, so that R_1, R_2, R_4 were grouped into module 1, and R_3 was grouped by itself into module 2, while S_2 corresponded to the partition {1,1,2,2}. We fixed the FC state sequence such that each individual spent half of the time in S_1 and then transitioned to S_2, i.e.

$$H_t = \begin{cases} S_1, & t = 1, \ldots, 180 \\ S_2 & t = 181, \ldots, 360 \end{cases}. \tag{4.6}$$

Thus, for the first 180 TRs, regions R_1, R_2, R_4 are connected and R_4 is unconnected, while for the last 180 TRs, regions R_1, R_2 are connected, as are regions R_3, R_4. This allowed for

the comparison of dFC between three types of region pairs: connected (within-module e.g. R_1-R_2), unconnected (between-module e.g. R_1-R_3), and a dynamic connection (within-module to between-module e.g. R_1-R_4). We then generated fMRI-like data using the simulation framework described above.

4.2.2 Base Simulation 2 (32 ROIs)

In this setting, we generated a total of $N_{states} = 9$ FC states, each consisting of a partition of the 32 ROIs into exactly 5 modules. For each FC state we generated a module label for each region from the numbers $\{1, \ldots, 5\}$ uniformly. If a FC state did not contain all 5 modules, we repeated this process. To ensure that no two FC states were too similar to each other, we computed the normalised mutual information (NMI) between each pair of state vectors, repeating the whole process if the maximum pairwise NMI exceeded 0.5. For each individual, we generated a random sequence of FC states under the assumption that a brain remained in a FC state for a fixed period of time before switching to any other FC state. Each FC state thus lasted a quarter of the total time period if three FC state transitions were specified, or half of the period if just one FC state transition was specified. We then generated fMRI-like data using the simulation framework described above.

4.3 Methods: specific simulations

For the first four simulations described here, we used Base Simulation 1 to generate the data. To measure dFC, we applied a sliding-window analysis. We used a tapered-cosine (Tukey) window (see Equation 4.2 of width $\omega = 30$ TRs with a total taper section of length 15 TRs ($\alpha = 0.5$). We slid the windows $\tau = 1$ time point at each step, yielding a total of 331 windows. For each window w and for each pair of regions R_i, R_j, we first calculated the Pearson correlation $\rho^j_{R_1,R_2}$ between the corresponding (weighted) time series. We then applied a Fisher transformation

$$z^j_{R_1,R_2} = \frac{1}{2} \log \left(\frac{1 + \rho^j_{R_1,R_2}}{1 - \rho^j_{R_1,R_2}} \right). \tag{4.7}$$

The effect of the Fisher transformation is to stabilise the variance of the correlation estimate across different values of $\rho^j_{R_1,R_2}$. In particular, this facilitates the comparison of the

variability in the connectivity estimate across different connection types (e.g. connected v. unconnected).

We then computed the standard deviation $\sigma_{corr(R_i,R_j)}$ of the time series of correlation values between each pair of regions. This measure is commonly defined as a proxy for dynamic functional connectivity. We also used the variance and the interquartile range of the correlation time series as alternative measures of dFC but these did not produce materially different results. For each set of parameters, we simulated 100 replicates in order to account for the randomness inherent in the data generation and also to evaluate the variability of our measure of dFC.

To assess their impact on this measure of dFC, we plot the SD of the (Fisher-transformed) correlation against the range of parameter values. For a given parameter, we also perform a multiple regression for each pair of regions. That is, for a parameter ζ, we fit the model

$$\sigma_{corr(R_i,R_j)} = \alpha + \beta_1\zeta + \beta_2\zeta^2 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{4.8}$$

for each pair of distinct regions R_i and R_j. If the resulting 95% confidence intervals of the model coefficients β_1, β_2 do not cover zero, this indicates that the parameter ζ has an impact on estimated dFC.

4.3.1 Neural autocorrelation

To investigate the effect of varying neural autocorrelation on the analysis of dFC, we used Base Simulation 1. To control the neural autocorrelation, we varied the generation of the region-specific neural event sequences. We modelled the binary sequences as Markov chains dependent on two parameters: the equilibrium probability of an event P_{reg} and the lag-1 autocorrelation ρ_{reg}. Specifically, for region R_j, we let $\mathbb{P}(E_{t+1}^{(j)} = x | E_t^{(j)} = y) = P_{yx}$ for $t = 1, \ldots, T-1$. To start the chain at equilibrium, we set $\mathbb{P}(E_1^{(j)} = 1) = P_{reg}$. First, observe that, under equilibrium, we must have

$$P_{reg} = P_{11}P_{reg} + P_{01}(1 - P_{reg}) \tag{4.9}$$

Second, by noting that $\mathrm{Var}(E_t^{(j)}) = \mathrm{Var}(E_1^{(j)}) = P_{reg}(1 - P_{reg})$, we can compute the lag-1 autocorrelation ρ_{reg} in terms of the transition probability P_{11} and the equilibrium probability P_{reg}:

$$
\begin{aligned}
\rho_{reg} &:= \frac{\mathbb{E}\left[(E_{t+1}^{(j)} - P_{reg})(E_t^{(j)} - P_{reg})\right]}{\sqrt{\mathrm{Var}(E_t^{(j)})\mathrm{Var}(E_{t+1}^{(j)})}} \\
&= \frac{\mathbb{E}\left[E_{t+1}^{(j)} E_t^{(j)}\right] - P_{reg}^2}{P_{reg}(1 - P_{reg})} \\
&= \frac{\mathbb{E}\left[E_{t+1}^{(j)} E_t^{(j)} | E_t^{(j)} = 1\right] P_{reg} - P_{reg}^2}{P_{reg}(1 - P_{reg})} \\
&= \frac{P_{11} P_{reg} - P_{reg}^2}{P_{reg}(1 - P_{reg})} \\
&= \frac{P_{11} - P_{reg}}{1 - P_{reg}}
\end{aligned}
\tag{4.10}
$$

By rearranging this expression, we see that $P_{11} = P_{reg} + (1 - P_{reg})\rho_{reg}$. Then, by (4.9), we find that $P_{01} = (1 - P_{reg})\rho_{reg}$. Since $P_{00} = 1 - P_{01}$ and $P_{10} = 1 - P_{11}$, this fully specifies the transition matrix P, allowing us to generate event sequences with the desired autocorrelation ρ_{reg} and equilibrium probability P_{reg}.

The default value in Base Simulation 1 is the special case of $P_{reg} = 0.5$ and $\rho_{reg} = 0$, indicating no autocorrelation, and is the value used in later simulations. For the purposes of this simulation, we kept the equilibrium probability of an event fixed at $P_{reg} = 0.5$, thus ensuring that the expected number of events was constant at 180. We generated data for $\rho_{reg} = -0.8, -0.7, \ldots, 0.8$ with the remainder of the simulation following the simulation framework described in Section 4.2. We also performed this analysis with a range of values of the other parameters: $P_{mod} = (0.25, 0.5, 0.75)$, $P_{reg} = (0.25, 0.5, 0.75)$, $a_{mod} = (1, 2, 3)$, and $\sigma_{noise} = (0.2, 0.5, 1)$.

Note that this region-specific signal can be considered a source of noise (as opposed to the module-specific events that drive the connectivity "signals"). The autocorrelation in this neural noise contributes to the temporal autocorrelation observed in the fMRI time series, which, once combined with the white noise measurement noise below, produces the "AR(1)+white noise" that characterises fMRI noise (at least after high-pass filtering; [48]). Nonetheless, in real fMRI data, there are other sources of coloured noise, such as those induced by respiratory and cardiac signals, and by head-movement (see e.g. [163]), which could also differ across groups.

A common procedure to correct for temporal autocorrelation in fMRI data is that of

prewhitening. While the impact of prewhitening on static functional connectivity has already been studied [11], the effect of prewhitening on dynamic connectivity is not as well evidenced. To assess its impact on estimated dFC, we performed the same analysis with a prewhitening step before the sliding-window procedure. To prewhiten the data, we used the strategy outlined in a previous study [51]. Briefly, we modelled the autocorrelation by a weighted sum of 8 exponential random variables with half-lives ranging from 0.5 to 64 TRs. The weights were then estimated via a restricted maximum likelihood approach. Finally, the resulting model fit was inverted to prewhiten the time series.

We do not perform prewhitening in the remaining analyses for the following reasons. Firstly, prewhitening is actually rarely done for resting-state connectivity analyses (see [11]). Secondly, our prewhitening requires efficient estimation of the temporal autocorrelation of the data. In real data, the efficiency of estimating the temporal autocorrelation matrix can be improved by pooling across voxels (within an ROI for example). Because we only simulate single time-series for each region in our model, our prewhitening will be inefficient (noisy). Thirdly, our prewhitening assumes stationarity of the time-series autocorrelation, and to estimate dynamic (window-specific) autocorrelation would be noisy and even less efficient. A full investigation of the effect of prewhitening on dynamic functional connectivity is beyond the scope of this work, although this topic deserves a more in-depth study.

4.3.2 Haemodynamic response function

To demonstrate the impact of the HRF on dFC, we generated data with various HRFs, two of which are shown in Figure 4.2. We varied two of the HRF parameters, the dispersion of peak response p_3 (the width of the initial peak) and the delay of response p_1, while the other 5 HRF parameters were held constant at their default values (see Table 4.1). In this simulation, we generated data using each HRF with peak dispersion $p_3 := \sigma_{HRF} = 0.6, 0.8, \ldots, 2.4$ and response delay $p_1 := \tau_{HRF} = 5, 5.2, \ldots, 9$s.

4.3.3 Connectivity strength

In our simulation framework, neural noise corresponds to the neural events that do not contribute to the connectivity structure. In other words, neural noise corresponds to the region-specific neural events, while the module-specific events are those which drive the connectivity structure. To investigate how connectivity strength affects dFC estimation,

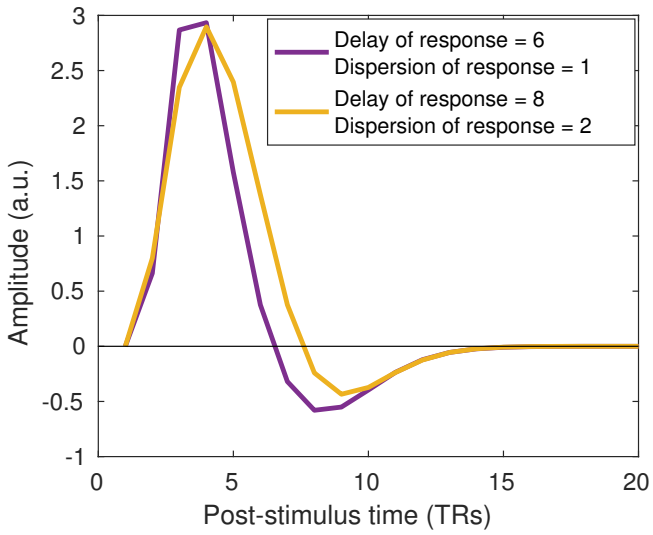


Figure 4.2: Two example haemodynamic response functions (HRF) based on different values for the dispersion of response and the delay of response. The purple line represents the default HRF used in the Base Simulations. We used a sampling rate of TR = 2s.

we varied the amplitude a_{mod} of the module-specific events relative to the region-specific events, which we fix to have amplitude $a_{reg} = 1$. In this simulation, we generated data for $a_{mod} = 0.5, 1, \ldots, 5$. We also performed this analysis with a range of values of the other parameters: $P_{mod} = (0.25, 0.5, 0.75)$, $P_{reg} = (0.25, 0.5, 0.75)$, $\rho_{reg} = (-0.5, 0, 0.5)$, and $\sigma_{noise} = (0.2, 0.5, 1)$.

4.3.4 Measurement noise

To investigate how measurement noise affects dFC estimation, we varied the amount of white noise added to the fMRI-like time series. We generated data with white noise of standard deviation $\sigma_{noise} = 0, 0.1, \ldots, 2.5$ times that of the signal. We also performed this analysis with a range of values of the other parameters: $P_{mod} = (0.25, 0.5, 0.75)$, $P_{reg} = (0.25, 0.5, 0.75)$, $\rho_{reg} = (-0.5, 0, 0.5)$, and $a_{mod} = (1, 2, 3)$.

4.3.5 k-means

A key assumption in many FC state-based methods is that there is a common set of FC states across individuals. In particular, it is assumed that, at rest, different participants transition within a comparable set of brain states. The k-means method for characterising dynamic functional connectivity aims to identify the common FC states by performing a

k-means clustering on windowed correlation matrices [7]. The k-means algorithm achieves this by partitioning the set of correlation matrices into k clusters such that each matrix belongs to the cluster with the nearest mean. Each cluster is thought to reflect a distinct brain state. State transitions for a given individual can then be identified through changes in the cluster labellings across consecutive windows in the brain scan time series.

In this simulation, we investigated what happens when the assumption of a common set of FC states does not hold. Specifically, we simulated 32-ROI fMRI data, using Base Simulation 2, from two groups of 100 people: Group 1 could visit 9 distinct FC states, and Group 2 could only visit 6 of these 9 FC states. In other words, the number of FC states $N_{states}^{(1)}$ for Group 1 is $N_{states}^{(1)} = 9$ but $N_{states}^{(1)} = 6$ for Group 2. Individuals in both groups experienced 3 FC state transitions in the course of a scan. FC state sequences for both groups were thus of length 4, with sequences for group 2 restricted to the FC states $\{1, \ldots, 6\}$. For Group 2, individuals were equally likely to be in each of the first six FC states. For group 1, in contrast, individuals were twice as likely to be in FC states 7,8 and 9 than in the first six FC states. The remainder of the simulation for each individual then followed the generic simulation framework.

We estimated correlation matrices for each position of the sliding-window analysis, as in the previous simulations, to produce 331 (Fisher-transformed) correlation matrices of size 32×32. The upper triangular part of this correlation matrix was vectorised to yield 331 correlation vectors of length 496 per subject. The k-means clustering was then performed on the set of all these vectors, pooled across subjects, with the ℓ_2 norm as distance measure. Centroids were initialised using the k-means++ algorithm in Matlab and analysis was repeated 40 times with different initial centroids to avoid sub-optimal clusterings. We investigated the performance of the clustering with $k = 1, \ldots, 12$ for the two groups separately and combined.

For each k, the algorithm returned a sequence of FC state labels for each subject, and the centroid of each of the k FC states. As the recovered FC state labels are arbitrary, the labels do not necessarily match those of the true FC states. While one can permute the recovered FC state labels to maximise the overlap with the true FC state sequence, this becomes more difficult with multiple subjects. Additionally, a simple relabelling does not take into account that incorrectly labelled FC states are not equally wrong. For example, if in two distinct windows, a subject is in FC state 1, but a k-means analysis recovers FC states 2 and 3 respectively (after relabelling), it may be the case that FC state 2 is closer to FC state 1 than FC state 3.

To circumvent the mislabelling and to enable comparisons of performance across different values of k, we replaced FC state labels by correlation matrices. For the recovered FC state sequences, we used the corresponding FC state centroid as calculated by the k-means algorithm. For the true sequences, we replaced a FC state label by a 'true' correlation matrix for that state. Recall that, in the 360 TRs simulated, a brain experienced 3 FC state transitions so that each FC state lasted 90 TRs. For each FC state, we first calculated the correlation matrix for each window of width 90 TRs in which all time points are in that FC state. We then took the 'true' correlation matrix as the average of all the corresponding correlation matrices in the same FC state across all subjects.

At each time point, we then computed the centroid error as the ℓ_2 distance from the 'true' correlation matrix to the centroid of the recovered FC state at that time. We thus used two measures of performance: the average number of detected FC state changes across subjects, and the mean centroid error across all time points and subjects.

Note that, in typical task-free fMRI analyses, we do not know the ground truth. In our simulation, however, we assumed that the states of all participants were drawn from a larger common pool of states. The k-means algorithm identifies comparable connectivity patterns and groups them into states. It does not take into account the order in which the states occur for an individual and so a state can occur at different times for different participants. This allowed us to cluster across individuals and time points.

4.3.6 Multilayer modularity

The multilayer modularity approach [15, 16] characterises the dynamic connectivity structure for an individual as a single multilayer network, with nodes corresponding to brain regions in different windows, or "layers". A community detection algorithm is then used to identify modules of connected regions at different points during the scan. Changes in connectivity for each region are then associated with changes in module assignment.

In many studies of dFC, a question of interest is whether groups differ in the degree to which connections between regions are static versus dynamic. Here, we examined how the frequency of state transitions could be detected using a multilayer modularity algorithm, and how the choice of parameters affected these results. To this end, we simulated 32-ROI data, using Base Simulation 2, from two groups of 50 people: individuals in group 1 experienced 3 FC state transitions, and group 2 experienced just 1 FC state transition in the course of a scan. FC state sequences for group 1 were thus of length 4, while FC state sequences for group 2 were of length 2. All individuals could visit the same 9 FC

states. The remainder of the simulation for each individual then followed the simulation framework described in Section 4.2. We applied the same sliding-window analysis as in the previous simulations, again using a window of width $\omega = 30$, calculating pairwise (Fisher-transformed) Pearson correlation for each window and for each pair of regions. In this case, however, we slid windows in steps of $\tau = 30$ TRs (instead of $\tau = 1$ TR) resulting 12 non-overlapping windows of width $\omega = 30$ TRs. This is based on a previous application of the multilayer modularity approach to fMRI data [15].

The non-overlapping sliding-window analysis yields a correlation matrix A with A_{ijl} corresponding to the correlation between regions R_i and R_j in window l. For each partitioning of regions into modules, the following multilayer modularity index is defined as a measure of the quality of the partition:

$$Q = \frac{1}{2\mu} \sum_{i=1}^{N_{reg}} \sum_{j=1}^{N_{reg}} \sum_{l=1}^{N_{win}} \sum_{r=1}^{N_{win}} \left[\left(A_{ijl} - \gamma \frac{k_{il} k_{jl}}{2m_l} \right) \delta_{l,r} + \omega \delta_{i,j} \delta_{l-r,1} \right] \delta(g_{il}, g_{jr}),$$

where γ and ω are hyperparameters, N_{reg} is the number of regions, N_{win} is the number of windows, g_{il} is the module assignment of region R_i in window l, $k_{il} = \sum_j A_{ijl}$, $2m_l = \sum_{ij} A_{ijl}$, and $2\mu = \sum_{jl}(k_{jl} + \sum_r \omega \delta_{l-r,1})$. The δ function is defined such that $\delta_{i,j} \equiv \delta(i,j) = 1$ if $i = j$ and is equal to 0 otherwise. In our simulation, $N_{reg} = 32$ and $N_{win} = 12$. The regions can then be partitioned into modules by attempting to maximise the modularity index Q using a generalised Louvain algorithm [100].

We investigated the effect of varying the hyperparameters γ and ω on the accuracy of the subsequent partitions. Broadly speaking, γ controls the resolution of the partitioning within layers so that a high value of γ encourages regions to be grouped into smaller modules, thus increasing the total number of modules. On the other hand, ω influences the 'stickiness' of module assignments between consecutive layers. Thus, a high value of ω encourages fewer module changes for an individual region.

To assess the performance of the algorithm, we used two error measures. Firstly, we created "incidence matrices" of size $32 \times 32 \times 12$, which contained an entry of 1 if the corresponding pair of regions had the same module label during a given window, and 0 otherwise. We then calculated the error in connectivity structure as the Hamming distance of the recovered incidence matrix (based on the multilayer modularity partitioning) to the true incidence matrix (based on the original module structure) at that time. The Hamming distance between two matrices of the same size is given by the number of elements at which the matrices differ. Secondly, we computed the mean flexibility for

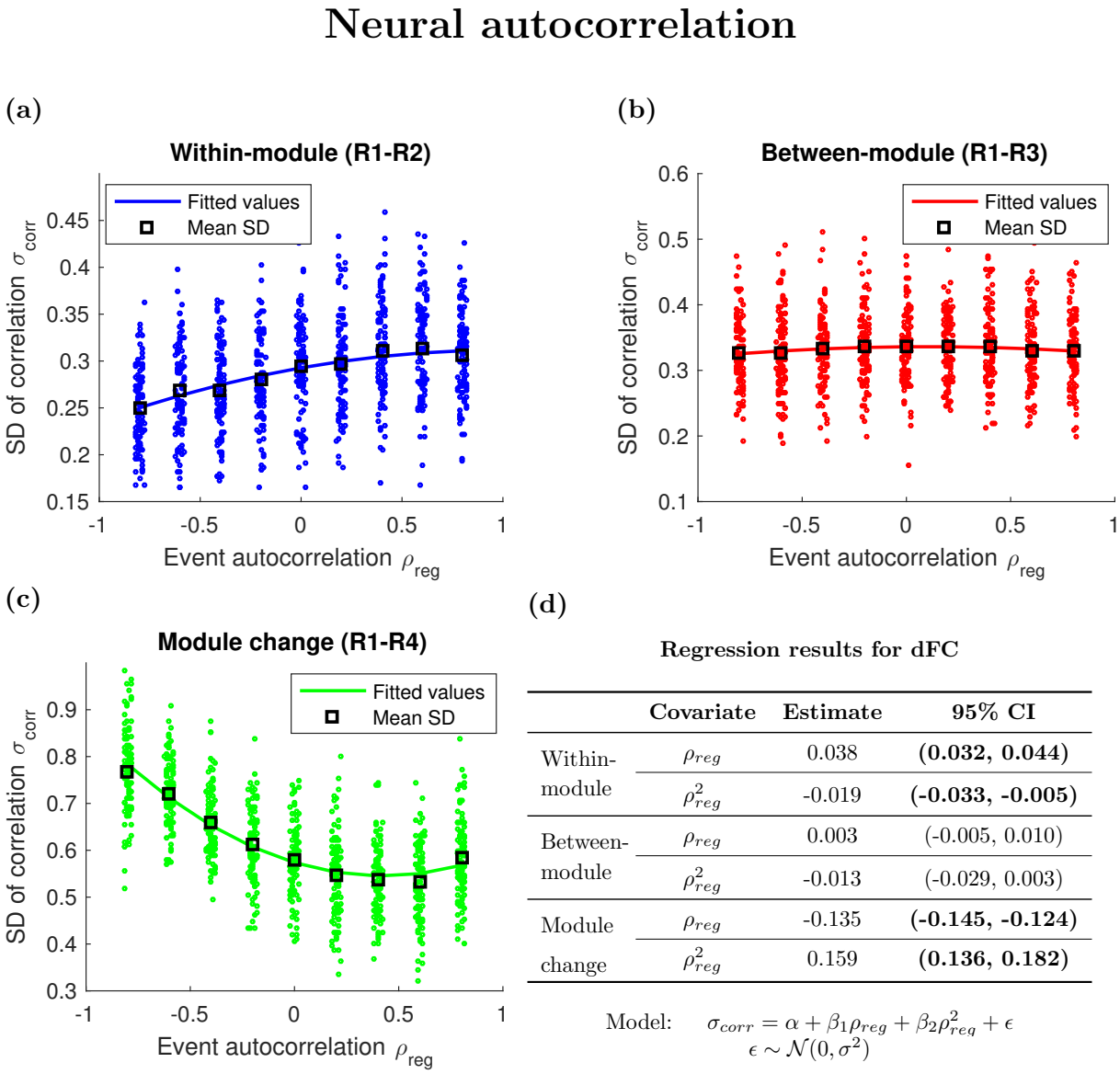
each subject. The flexibility of a region is calculated by the number of times the region changes module assignment divided by the total possible number of module changes [15]. The mean flexibility is then given by the mean region flexibility over all 32 regions. Note that the expected mean flexibility for a young individual in our simulation is approximately 4/55 (1 state change out of a possible 11 with the probability of a region changing module at a state change of approximately 0.8), compared to 12/55 for an old individual. For each subject, we ran the algorithm for $\gamma = 0.75, 1, \ldots, 2.5$ and $\omega = 0.25, 0.5, \ldots, 4.5$.

4.4 Results

We first investigate whether unaccounted heterogeneity can impact estimates of dFC. Here we examine four sources of heterogeneity: 1) individual differences in neural autocorrelation, 2) shape of the HRF, 3) connectivity strength and 4) measurement noise. We consider a simple, yet common, measure of dynamic connectivity, namely the standard deviation (SD) of correlation values across sliding windows. We calculated this measure for three types of true connectivity: 1) static, positive connections between regions within the same module, 2) static, zero connections between regions in different modules, and 3) dynamic changes between positive and zero connections when a region switched from being in the same module to being in a different module as another region (see Section 4.2). As expected, across all the simulations, estimated dFC for the dynamic connections is higher than the estimated dFC for both types of static connections.

4.4.1 Neural autocorrelation

In this simulation, we investigated the association between neural autocorrelation and estimated dFC, making sure that differences in neural autocorrelation were not associated with differences in 'true' dynamic connectivity. This was achieved by varying the autocorrelation ρ_{reg} of *region-specific* events but keeping the autocorrelation of the module-specific events fixed at zero. Recall that region-specific events are generated independently of the connectivity structure so, under our simulation framework, changing them should have no effect on dFC. Note that the underlying dFC structure is held constant across all iterations. We estimated dFC for three types of connection: a static, positive connection, a static zero connection, and a dynamic connection (from positive to zero) for $\rho_{reg} = -0.8, -0.6, \ldots, 0.8$.



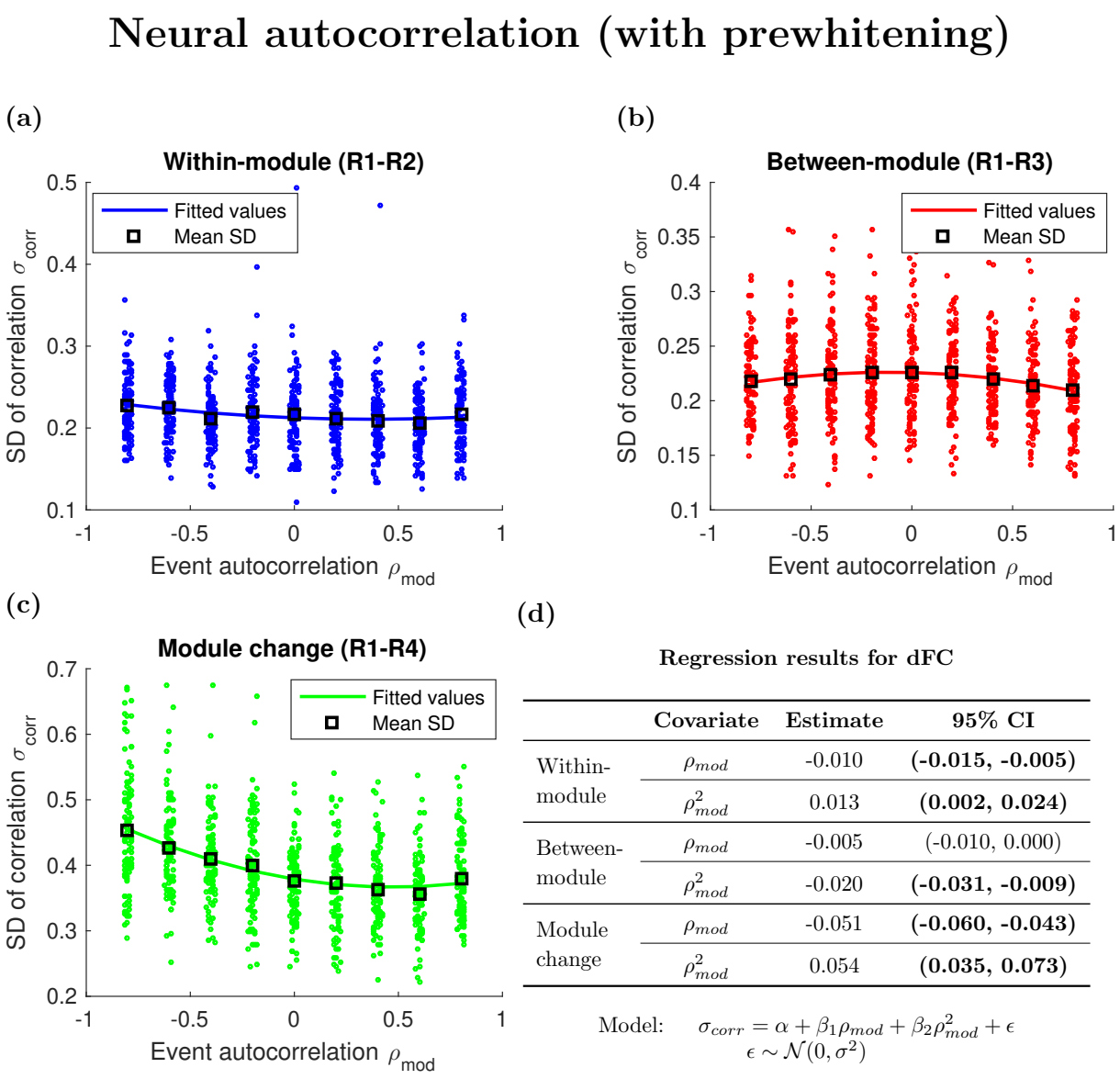
(d)

Regression results for dFC

	Covariate	Estimate	95% CI
Within-module	ρ_{reg}	0.038	**(0.032, 0.044)**
	ρ^2_{reg}	-0.019	**(-0.033, -0.005)**
Between-module	ρ_{reg}	0.003	(-0.005, 0.010)
	ρ^2_{reg}	-0.013	(-0.029, 0.003)
Module change	ρ_{reg}	-0.135	**(-0.145, -0.124)**
	ρ^2_{reg}	0.159	**(0.136, 0.182)**

Model: $\sigma_{corr} = \alpha + \beta_1 \rho_{reg} + \beta_2 \rho^2_{reg} + \epsilon$
$\epsilon \sim \mathcal{N}(0, \sigma^2)$

Figure 4.3: The impact of neural autocorrelation on estimated dFC, measured by the SD of the correlation time series, between (a) statically connected regions (within-module), (b) unconnected regions (between-module), and (c) dynamically connected regions (module change), with (d) the results of a multiple regression. Ri refers to Region i. Region pair R1-R4 has a dynamic connection so the true dFC should be higher than region pairs R1-R2 and R1-R3, which have a static connection. Estimated dFC between statically connected (within-module) regions increased with neural autocorrelation, while estimated dFC for dynamically connected regions (module change) decreased with increased neural autocorrelation. The neural autocorrelation is varied independently of the underlying connectivity structure, so changing it should have no effect on dFC. The multiple regression assesses the impact of neural autocorrelation on estimated dFC with a statistically significant effect indicated by a 95% confidence interval (CI) in bold type. The solid lines in (a-c) correspond to the fitted values of the multiple regression.



Regression results for dFC

	Covariate	Estimate	95% CI
Within-module	ρ_{mod}	-0.010	**(-0.015, -0.005)**
	ρ^2_{mod}	0.013	**(0.002, 0.024)**
Between-module	ρ_{mod}	-0.005	(-0.010, 0.000)
	ρ^2_{mod}	-0.020	**(-0.031, -0.009)**
Module change	ρ_{mod}	-0.051	**(-0.060, -0.043)**
	ρ^2_{mod}	0.054	**(0.035, 0.073)**

Model: $\sigma_{corr} = \alpha + \beta_1 \rho_{mod} + \beta_2 \rho^2_{mod} + \epsilon$
$\epsilon \sim \mathcal{N}(0, \sigma^2)$

Figure 4.4: The impact of neural autocorrelation on estimated dFC *after prewhitening*, measured by the SD of the correlation time series, between (a) statically connected regions (within-module), (b) unconnected regions (between-module), and (c) dynamically connected regions (module change), with (d) the results of a multiple regression. Ri refers to Region i. Region pair R1-R4 has a dynamic connection so the true dFC should be higher than region pairs R1-R2 and R1-R3, which have a static connection. Estimated dFC between statically connected (within-module) regions decreases slightly with neural autocorrelation, while estimated dFC for dynamically connected regions (module change) decreased with increased neural autocorrelation. The impact of neural autocorrelation on estimated dFC is dampened when prewhitening is performed compared to when the simulated time series are not prewhitened (compare to Figure 4.3). The neural autocorrelation is varied independently of the underlying connectivity structure, so changing it should have no effect on dFC. The multiple regression assesses the impact of neural autocorrelation on estimated dFC with a statistically significant effect indicated by a 95% confidence interval (CI) in bold type. The solid lines in (a-c) correspond to the fitted values of the multiple regression.

Figure 4.3 illustrates the impact of neural autocorrelation on estimated dFC. Although Figure 4.3b shows that estimated dFC for the unconnected regions remains unaffected by changes in neural autocorrelation, Figure 4.3a demonstrates that estimated dFC between the positively connected regions is higher as neural autocorrelation increases. In contrast, Figure 4.3c shows that estimated dFC decreased between the dynamically-connected regions as the autocorrelation increased. In other words, the ability to detect a difference in dFC between dynamically connected and statically connected regions decreases with higher levels of neural autocorrelation. These observations are supported by the multiple regression in Figure 4.3d: neural autocorrelation had a statistically significant effect on estimated dFC for the positively connected and dynamically connected regions, but not for the unconnected regions. These results suggest that observed dFC may vary substantially due to differences in neural autocorrelation, even though the true dFC was identical across individuals.

To assess the effect of prewhitening, we carried out the same analysis with a prewhitening step before performing the sliding-window procedure (see Section 4.3.1). Figure 4.4 shows that neural autocorrelation impacts estimated dFC even after prewhitening. 4.4c shows that although prewhitening attenuates the negative effect of increasing neural autocorrelation on dFC estimates for truly dynamic connections, a significant negative effect of neural autocorrelation remains after prewhitening. Furthermore, 4.4a shows that the positive effect of increasing neural autocorrelation on dFC estimates for within-module static connections actually becomes a significant negative effect after prewhitening. Figure 4.4b shows that for the unconnected regions, estimated dFC increases as neural autocorrelation increases from -0.8 to 0, but decreases from 0 to 0.8. Thus prewhitening does not fully solve the problem.

We also performed the analysis (without prewhitening) for a range of values for the other simulation parameters, namely $P_{mod} = 0.25, 0.5, 0.75$, $P_{reg} = 0.25, 0.5, 0.75$, $a_{mod} = 1, 2, 3$, and $\sigma_{noise} = 0.2, 0.5, 1$. Neural autocorrelation impacted estimated dFC between dynamically connected regions for the whole explored parameter space while the impact of neural autocorrelation on estimated dFC between the statically connected regions for all the parameter values differed throughout the parameter space.

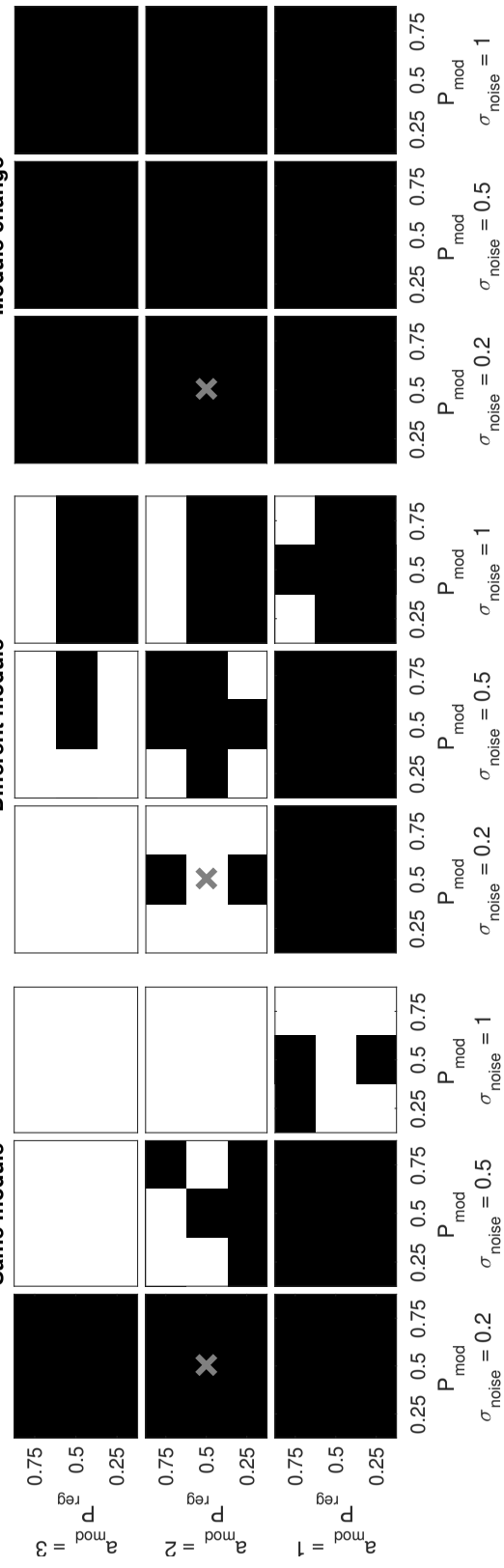


Figure 4.5: The effect of neural autocorrelation on estimated dFC for a range of parameter values. To assess the impact of neural autocorrelation ρ_{reg} on estimated dFC, σ_{corr}, we fit the linear model: $\sigma_{corr} = \alpha + \beta_1 \rho_{reg} + \beta_2 \rho_{reg}^2 + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The statistical significance of the effect of ρ_{reg} or ρ_{reg}^2 on σ_{corr} is determined by whether the 95% confidence interval for the corresponding coefficient contains zero. Thus neural autocorrelation has a statistically significant impact on estimated dFC if at least one of the confidence intervals for ρ_{reg} or ρ_{reg}^2 does not contain zero. A statistically significant effect is indicated by a black square. We considered three values each of the equilibrium probability of module-specific events P_{mod}, the equilibrium probability of region-specific events P_{reg}, the autocorrelation of region-specific events ρ_{reg}, and the standard deviation of white noise σ_{noise}. A grey cross indicates the parameter values used in the main analysis (see Figure 4.3).

Regression results for dFC

	Within-module		Between-module	
Covariate	**Estimate**	**95% CI**	**Estimate**	**95% CI**
τ_{HRF}	0.093	**(0.084, 0.101)**	0.098	**(0.088, 0.107)**
σ_{HRF}	0.013	**(0.002, 0.024)**	0.032	**(0.020, 0.045)**
$\tau_{HRF} : \sigma_{HRF}$	0.000	(-0.001, 0.001)	-0.002	**(-0.004, -0.001)**
τ^2_{HRF}	-0.006	**(-0.006, -0.005)**	-0.006	**(-0.007, -0.005)**
σ^2_{HRF}	-0.004	**(-0.007, -0.001)**	-0.005	**(-0.008, -0.002)**

	Module change	
Covariate	**Estimate**	**95% CI**
τ_{HRF}	0.039	**(0.027, 0.052)**
σ_{HRF}	0.004	(-0.013, 0.021)
$\tau_{HRF} : \sigma_{HRF}$	-0.000	(-0.002, 0.001)
τ^2_{HRF}	-0.003	**(-0.004, -0.002)**
σ^2_{HRF}	-0.000	(-0.004, 0.004)

Model: $\sigma_{corr} = \alpha + \beta_1\tau_{HRF} + \beta_2\sigma_{HRF} + \beta_3\tau_{HRF}\sigma_{HRF} + \beta_4\tau^2_{HRF} + \beta_5\sigma^2_{HRF} + \epsilon$
$\epsilon \sim \mathcal{N}(0, \sigma^2)$

Table 4.2: Multiple regression results summarising the impact of the two HRF parameters on estimated dFC with a statistically significant effect indicated by a 95% confidence interval (CI) in bold type.

4.4.2 Haemodynamic response function

It has been shown that the haemodynamic response function (HRF) varies between different ages [73, 3, 39] and disease states [69]. To illustrate the effect of the HRF on observed dFC, we altered the HRF by varying the dispersion of the response σ_{HRF} and the delay of response τ_{HRF}. Note that the underlying dFC structure is held constant across all iterations. To measure dFC, we calculated the mean SD of the sliding-window correlation time series for the three types of region pairs (positively connected, unconnected, and dynamically connected) for $\sigma_{HRF} = 0.6, 0.8, \ldots, 2.4$ and $\tau_{HRF} = 5, 5.2, \ldots, 9$s.

Figures 4.6a and 4.6b show that a more temporally dispersed HRF (as often observed for older individuals) resulted in increased dFC between regions, when in truth the connectivity remained constant. Increases in both the dispersion of peak response σ_{HRF} and the delay of response τ_{HRF} resulted in increased estimated dFC. Figure 4.6c shows a similar effect for regions with a dynamic connection though, in this case, the increase was not as marked. Figure 4.6d compares the observed dFC for two of these HRFs. Individuals in Group 1 had a HRF with peak dispersion 1 and a response delay of 6s (purple

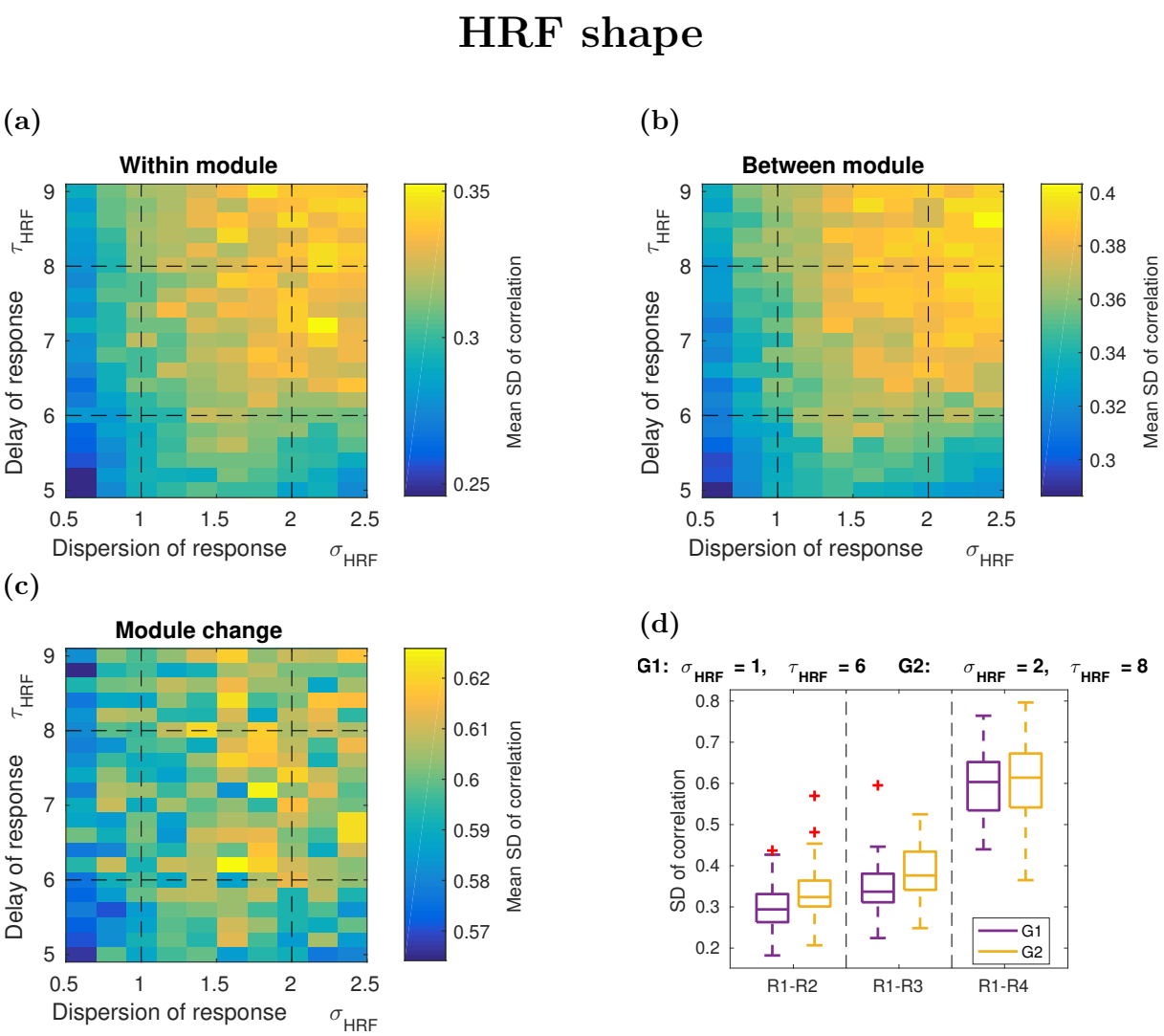


Figure 4.6: The impact of HRF shape on estimated dFC, measured by the SD of the correlation time series, between (a) statically connected regions (within-module), (b) unconnected regions (between-module), and (c) dynamically connected regions (module change), with (d) a comparison of estimated dFC for two groups and (e) a summary of results. Ri refers to Region i. The HRF was altered by varying two parameters: the dispersion of response σ_{HRF} and the delay of response τ_{HRF}. Individuals in G1 had an HRF with $\sigma_{HRF} = 1$ and $\tau_{HRF} = 6$ while individuals in G2 had an HRF with $\sigma_{HRF} = 2$ and $\tau_{HRF} = 8$. Each square in (a-c) corresponds to the mean (across individuals) estimated dFC for a pair of parameter values of $(\sigma_{HRF}, \tau_{HRF})$, with yellow indicating higher dFC. The dashed lines in (a-c) indicate the parameter values of the two groups compared in (d), whose HRFs are shown in Figure 4.2. A more dispersed HRF increased estimated dFC between all three types of region pairs, despite all individuals having identical dFC structure. We observed higher variability in dFC for the statically connected regions (within-module) as the dispersion and delay of the response increased. We saw a similar effect for the dynamically connected regions (module change), though this was less pronounced. The box plots in (d) illustrate a single comparison of two groups.

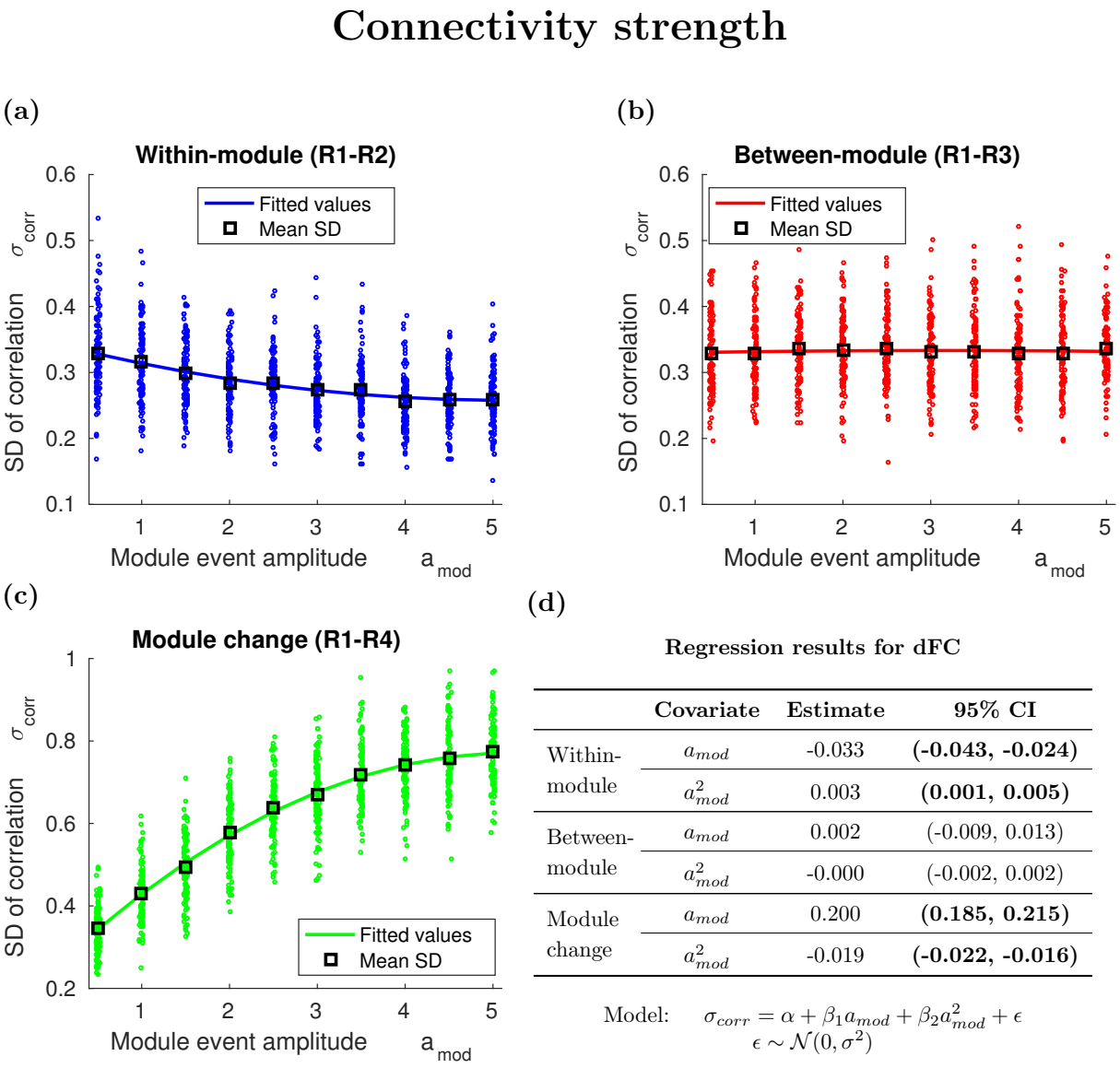
line in Figure 4.2), which might represent a younger sample. In contrast, individuals in Group 2 had a HRF with peak dispersion 2 and a response delay of 8s (yellow line in Figure 4.2), which might represent an older sample. For the static connections we see that, even though the true dFC was the same across groups, the observed dFC varied substantially between groups due to the shape of the HRF. These observations are supported by the multiple regression results in Table 4.2: both the dispersion of peak response σ_{HRF} and the delay of response τ_{HRF} had statistically significant effects on estimated dFC for the statically connected regions, but only the delay of response τ_{HRF} had a statistically significant impact for the dynamically connected regions.

4.4.3 Connectivity strength

In our simulation framework, connectivity strength corresponds to the amplitude of the module-specific events, since these drive the connectivity structure. In this simulation, we investigated the association between connectivity strength and estimated dFC. This was done by varying the amplitude of module-specific events a_{mod} while keeping the amplitude of the region-specific events fixed (i.e. the size of the connectivity "signal" versus region-specific neural "noise"). Note that the underlying dFC structure is held constant across all iterations. To measure dFC, we calculated the mean SD of the sliding-window correlation time series for the three types of region pairs (positively connected, unconnected, and dynamically connected) for $a_{mod} = 0.5, 1, \ldots, 5$.

Figure 4.7 illustrates the impact of connectivity strength on estimated dFC. As expected, Figure 4.7b shows that estimated dFC for the unconnected regions remains unaffected by changes in connectivity strength. However, Figure 4.7a demonstrates that estimated dFC between the positively connected regions is moderately lower as connectivity strength increases. In contrast, Figure 4.7c shows that estimated dFC increased between the dynamically-connected regions as connectivity strength increased. In particular, the ability to detect a difference in dFC between dynamically connected and statically connected regions decreases with lower connectivity strength. These observations are supported by the multiple regression results in Figure 4.7d: connectivity strength had a statistically significant effect on estimated dFC for the positively connected and dynamically connected regions, but not for the unconnected regions.

We also performed the analysis for a range of values for the other simulation parameters, namely $P_{mod} = 0.25, 0.5, 0.75$, $P_{reg} = 0.25, 0.5, 0.75$, $\rho_{reg} = -0.5, 0, 0.5$, and $\sigma_{noise} = 0.2, 0.5, 1$. Figure 4.8 shows that connectivity strength had a statistically sig-



Regression results for dFC

	Covariate	Estimate	95% CI
Within-module	a_{mod}	-0.033	**(-0.043, -0.024)**
	a_{mod}^2	0.003	**(0.001, 0.005)**
Between-module	a_{mod}	0.002	(-0.009, 0.013)
	a_{mod}^2	-0.000	(-0.002, 0.002)
Module change	a_{mod}	0.200	**(0.185, 0.215)**
	a_{mod}^2	-0.019	**(-0.022, -0.016)**

Model: $\sigma_{corr} = \alpha + \beta_1 a_{mod} + \beta_2 a_{mod}^2 + \epsilon$
$\epsilon \sim \mathcal{N}(0, \sigma^2)$

Figure 4.7: The impact of connectivity strength on estimated dFC, measured by the SD of the correlation time series, between (a) statically connected regions (within-module), (b) unconnected regions (between-module), and (c) dynamically connected regions (module change), with (d) the results of a multiple regression. Ri refers to Region i. Region pair R1-R4 has a dynamic connection so the true dFC should be higher than region pairs R1-R2 and R1-R3, which have a static connection. Increased amplitude of module-specific events resulted in decreased observed dFC for the positive static connected pair (within-module) but increased dFC for the dynamically connected pair (module change). The effect on the unconnected pair (between-module) was small. The connectivity strength, corresponding to the amplitude of module-specific events, is varied independently of the underlying connectivity structure, so changing it should have no effect on dFC. The multiple regression assesses the impact of connectivity strength on estimated dFC with a statistically significant effect indicated by a 95% confidence interval (CI) in bold type. The solid lines in (a-c) correspond to the fitted values of the multiple regression.

nificant impact on estimated dFC between dynamically connected regions for all the parameter values investigated, but the effect on statically connected regions depends on the choice of parameters. Altogether, these results suggest that observed dFC may vary substantially due to differences in connectivity strength, even though the true dFC was identical across individuals.

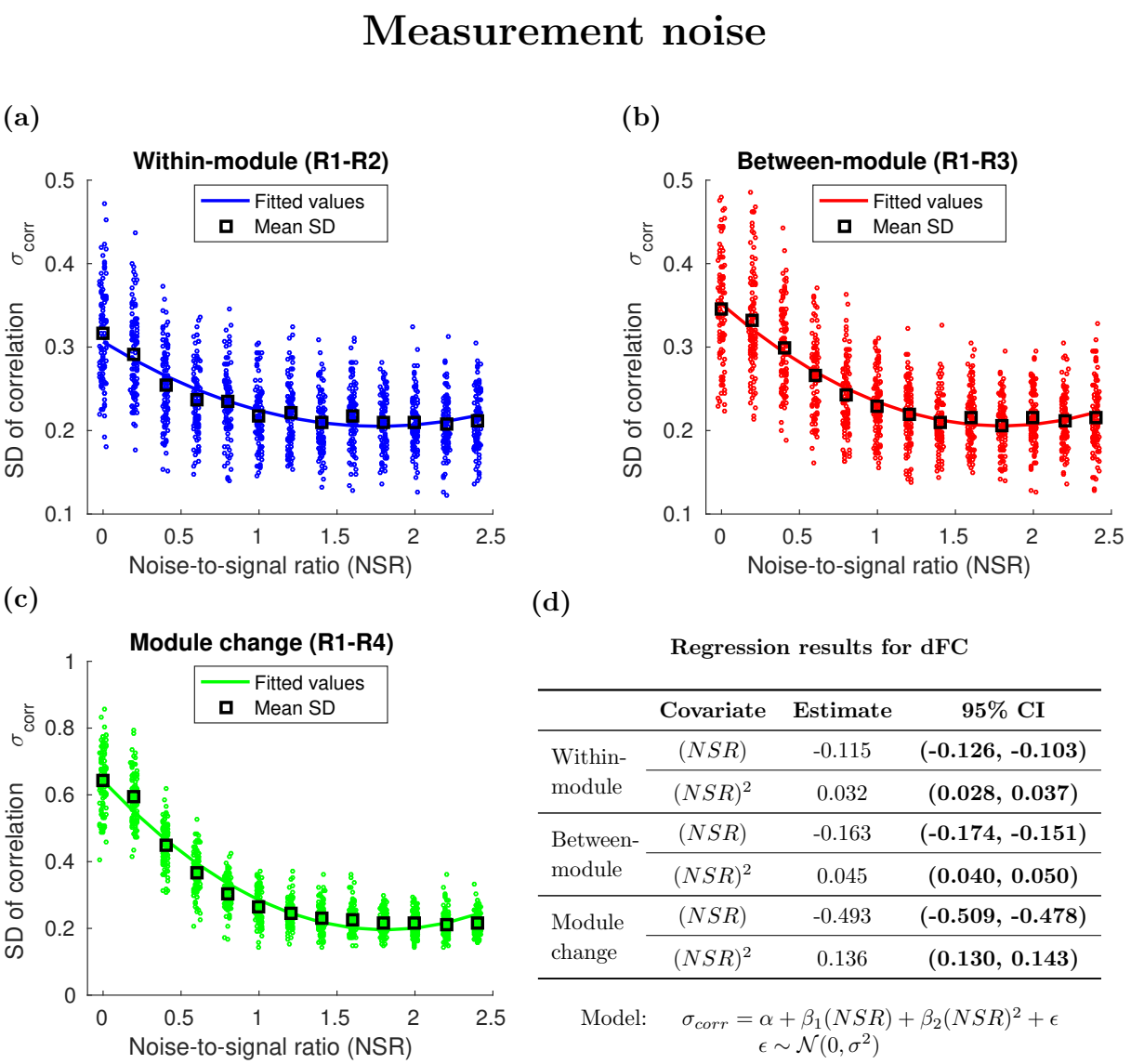
Figure 4.8: The effect of connectivity strength on estimated dFC for a range of parameter values. To assess the impact of connectivity strength a_{mod} on estimated dFC ,σ_{corr}, we fit the linear model: $\sigma_{corr} = \alpha + \beta_1 a_{mod} + \beta_2 a_{mod}^2 + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The statistical significance of the effect of a_{mod} or a_{mod}^2 on σ_{corr} is determined by whether the 95% confidence interval for the corresponding coefficient contains zero. Thus connectivity strength has a statistically significant impact on estimated dFC if at least one of the confidence intervals for a_{mod} or a_{mod}^2 does not contain zero. A statistically significant effect is indicated by a black square. We considered three values each of the equilibrium probability of module-specific events P_{mod}, the equilibrium probability of region-specific events P_{reg}, the autocorrelation of region-specific events ρ_{reg}, and the standard deviation of white noise σ_{noise}. The grey cross indicates the parameter values used in the main analysis (see Figure 4.7).

4.4.4 Measurement noise

We also investigated the effects of varying amounts of measurement noise on observed connectivity dynamics by generating data with white noise of SD $\sigma_{noise} = 0, 0.1, \ldots, 2.5$. Recall that the signal was rescaled to have SD 1, resulting in noise-to-signal ratios equal to σ_{noise} (ignoring neural noise).

Figure 4.9 shows the effect of varying measurement noise on the standard deviation of correlation. Figures 4.9a, 4.9b, and 4.9c show that for all three types of region pair, increasing the amount of measurement noise resulted in decreased observed dFC. This is supported by the multiple regression results in Figure 4.9d: measurement noise had a statistically significant effect on estimated dFC for all three connection types. Thus, noisier data resulted in lower estimated dFC even for the pair of ROIs that experienced a true change in connectivity. When the amplitude reached a certain threshold ($\sigma_{noise} > 2.0$), the white noise dominated the underlying fMRI signal, resulting in a mean dFC of 0.2 for all three types of connectivity.

Figure 4.10 shows the results of the same analysis for different combinations of values for the other parameters, namely $P_{mod} = 0.25, 0.5, 0.75$, $P_{reg} = 0.25, 0.5, 0.75$, $\rho_{reg} = -0.5, 0, 0.5$, and $a_{mod} = 1, 2, 3$. Here, measurement noise had a significant effect on estimated dFC between all three connectivity types for the whole explored parameter space, indicating that observed dFC may vary due to differences in noise even when the underlying true dFC remains constant.



	Covariate	Estimate	95% CI
Within-module	(NSR)	-0.115	**(-0.126, -0.103)**
	$(NSR)^2$	0.032	**(0.028, 0.037)**
Between-module	(NSR)	-0.163	**(-0.174, -0.151)**
	$(NSR)^2$	0.045	**(0.040, 0.050)**
Module change	(NSR)	-0.493	**(-0.509, -0.478)**
	$(NSR)^2$	0.136	**(0.130, 0.143)**

Model: $\sigma_{corr} = \alpha + \beta_1(NSR) + \beta_2(NSR)^2 + \epsilon$
$\epsilon \sim \mathcal{N}(0, \sigma^2)$

Figure 4.9: The impact of measurement noise on estimated dFC, measured by the SD of the correlation time series, between (a) statically connected regions (within-module), (b) unconnected regions (between-module), and (c) dynamically connected regions (module change), with (d) the results of a multiple regression. Ri refers to Region i. Region pair R1-R4 has a dynamic connection so the true dFC should be higher than region pairs R1-R2 and R1-R3, which have a static connection. Increased measurement noise resulted in decreased observed dFC for the three types of region pairs. This effect was particularly pronounced for the dynamically connected regions (module change). The multiple regression assesses the impact of measurement noise on estimated dFC with a statistically significant effect indicated by a 95% confidence interval (CI) in bold type. The solid lines in (a-c) correspond to the fitted values of the multiple regression.

Figure 4.10: The effect of measurement noise on estimated dFC for a range of parameter values. To assess the impact of measurement noise σ_{noise} on estimated dFC ,σ_{corr}, we fit the linear model: $\sigma_{corr} = \alpha + \beta_1\sigma_{noise} + \beta_2\sigma^2_{noise} + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The statistical significance of the effect of σ_{noise} or σ^2_{noise} on σ_{corr} is determined by whether the 95% confidence interval for the corresponding coefficient contains zero. Thus measurement noise has a statistically significant impact on estimated dFC if at least one of the confidence intervals for σ_{noise} or σ^2_{noise} does not contain zero. A statistically significant effect is indicated by a black square. We considered three values each of the equilibrium probability of module-specific events P_{mod}, the equilibrium probability of region-specific events P_{reg}, the autocorrelation of region-specific events ρ_{reg}, and the amplitude of module-specific events a_{mod}. A grey cross indicates the parameter values used in the main analysis (see Figure 4.9).

4.4.5 k-means

In the previous sections, we demonstrated that group differences in connectivity dynamics, as measured with simple sliding window approaches, can be due to other factors such neural autocorrelation. However, even when such differences do not exist, some common dFC methods may still detect artifactual group differences in dFC owing to unaccounted heterogeneity in the dynamic connectivity structure.

To investigate the effect of heterogeneity in the number of FC states attainable, we generated a set of data for two groups of 50 individuals: those in G1 could reach 9 FC states, and those in G2 could reach only 6 of these 9 FC states (see Section 4.3.5). If these FC states can be recovered accurately from the data, then a simple measure of dFC is the number of FC state transitions that occur. Importantly, in the simulations, the number of such transitions was identical across groups, namely three.

We used a k-means cluster analysis on the correlation matrices in an attempt to recover the underlying FC states. Figure 4.11 illustrates the performance of the k-means analysis for values of the hyperparameter $k = 1, \ldots, 12$. We ran the analysis for the two groups separately, and also for all 100 individuals together. Here, we perform the analysis with sliding-window width $w = 30$, though window widths $w = 60, 90, 120$ TRs, yielded broadly similar results (see Figures 4.12, 4.13 and 4.14). Further, Figure 4.15 shows that, when k is correctly estimated or only slightly misspecified, it becomes more difficult to estimate the states correctly as window length increases.

Figure 4.11a shows that when $k < 9$, the typical (combined) k-means analysis underestimated the number of FC state transitions for G1, while the number of FC state transitions for G2 was recovered more accurately. Unless the correct value of k was estimated, Figure 4.11c shows that the combined analysis leads to artifactual differences in dFC between groups. One might think that fitting the two groups separately would solve this problem. Figure 4.11b demonstrates a modest improvement in the error in number of FC state transitions for G1 (yellow line) when k was underestimated, but a steep increase in error for G2 when k was overestimated. This deterioration when $k > 6$ is to be expected because the k-means algorithm had to identify more FC states than are actually present in the G2. We see in Figure 4.11d that for the separate analysis, incorrect group differences were again found when k was overestimated or grossly misspecified for either group.

Figures 4.11e and 4.11f illustrate the differences between the groups for both analyses in mean centroid error, which is a measure of how well the complete space-time connec-

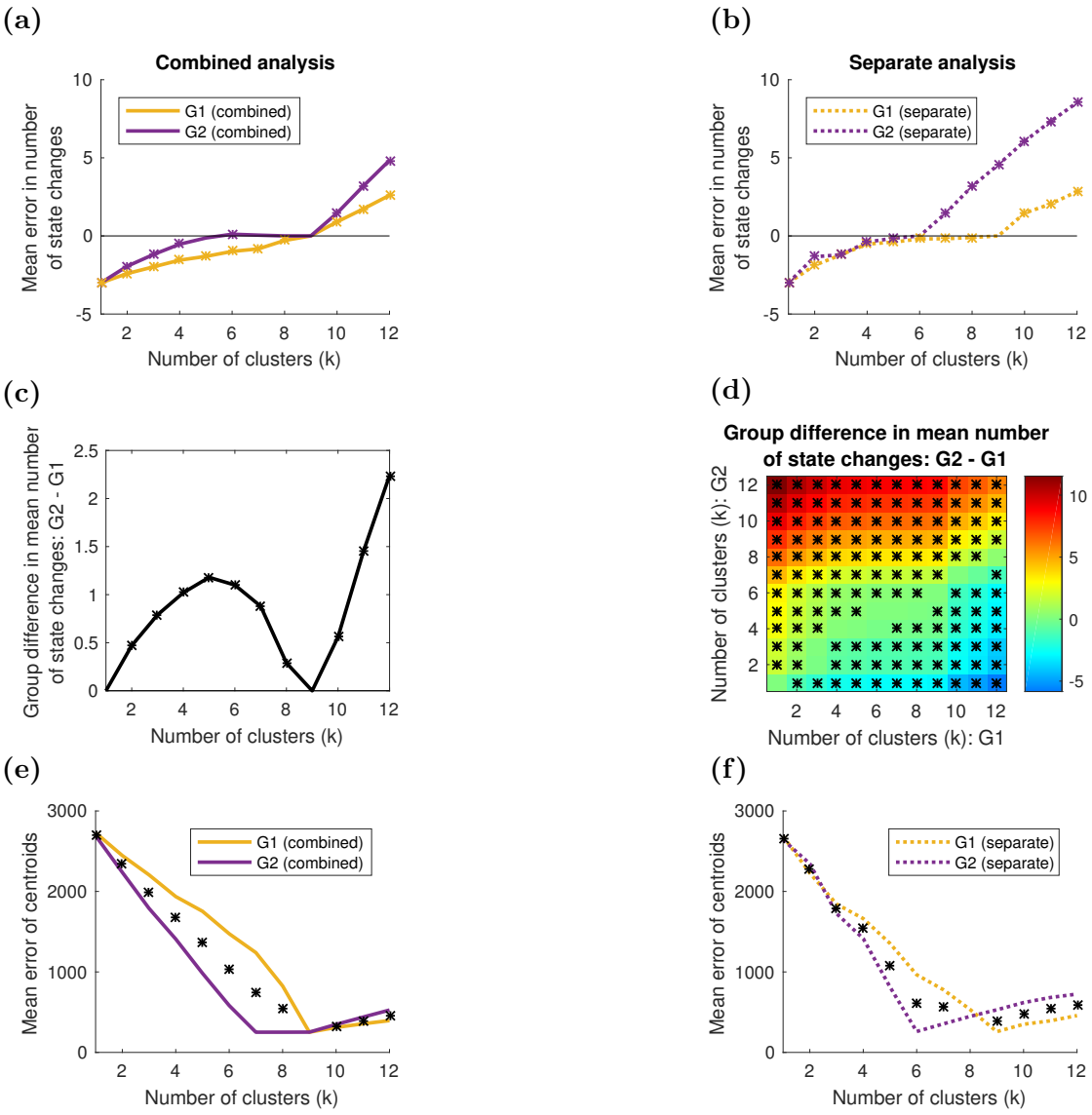


Figure 4.11: The results for the k-means simulation for window length $w = 30$. The true number of FC state transitions in this simulation is 3. Individuals in G1 could reach 9 FC states while individuals in G2 could reach only 6 of these 9 FC states. (a) When $k < 9$, the combined analyses underestimated the number of FC state transitions for G1 (solid yellow line), while the number of FC state transitions for G2 (solid purple line) was recovered more accurately. (b) The separate analysis showed an improvement in the error in number of FC state transitions for G1 (dotted yellow line) when k was underestimated. (c) Unless k was correctly estimated, the combined analysis yielded an incorrect group difference in number of FC state transitions. (d) For the separate analysis, incorrect group differences were only found when k was overestimated or grossly misspecified for either group. (e) In terms of recovered space-time connectivity structure, the combined analysis performed better for G2 (solid purple line) than G1 (solid yellow line) when k was underestimated. (f) If k was correctly specified or only slightly misspecified for either group, the separate analyses had a similar error in recovered connectivity structure. Asterisks in (a-d) indicate a statistically significant ($p < 0.05$) difference from zero, according to a Wilcoxon signed-rank test, while asterisks in (e-f) indicate a statistically significant ($p < 0.05$) difference between the two groups, according to a two-sample Wilcoxon rank-sum test.

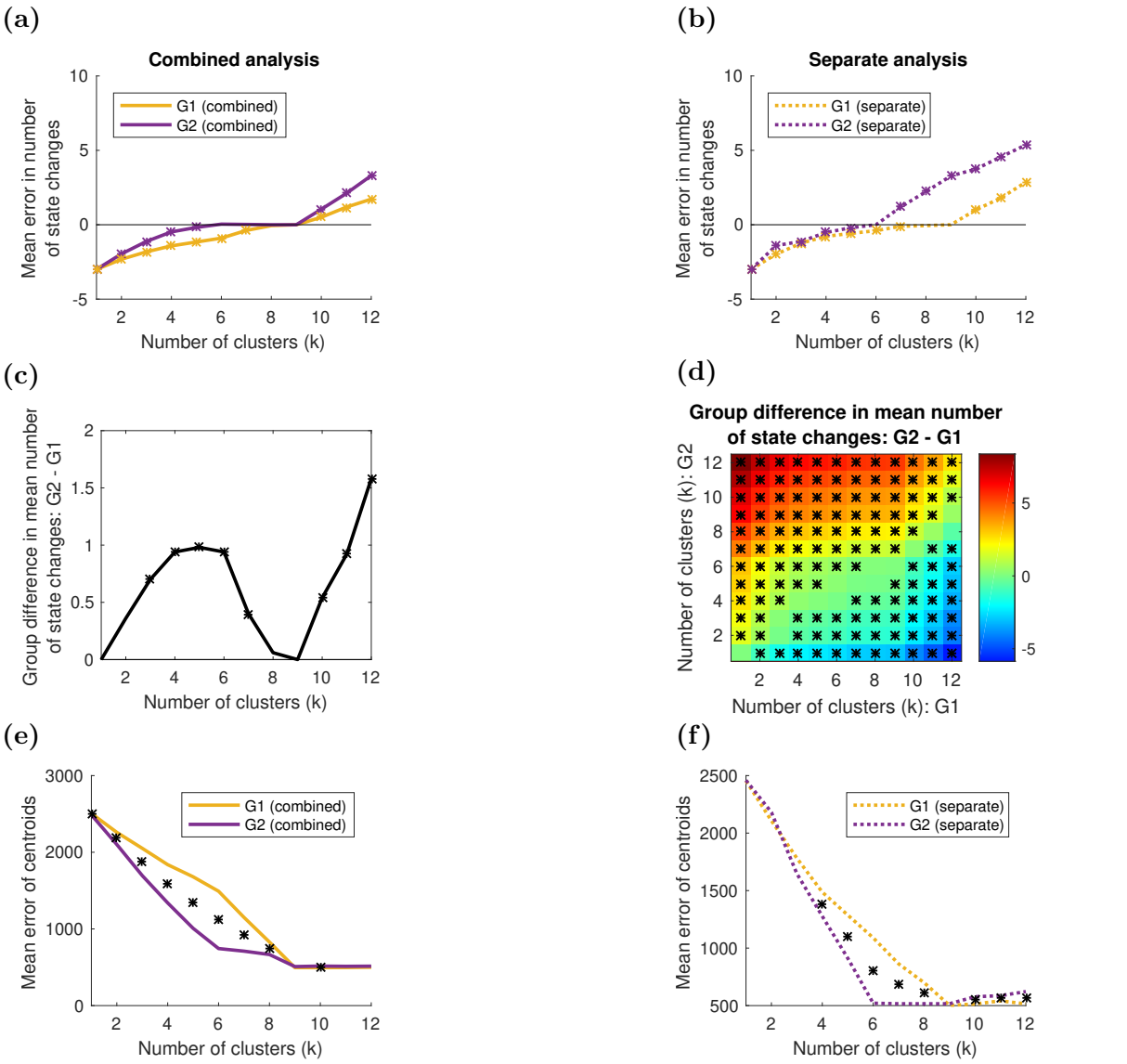


Figure 4.12: The results for the k-means simulation for window length $w = 60$. The true number of FC state transitions in this simulation is 3. Individuals in G1 could reach 9 FC states while individuals in G2 could reach only 6 of these 9 FC states. (a) When $k < 9$, the combined analyses underestimated the number of FC state transitions for G1 (solid yellow line), while the number of FC state transitions for G2 (solid purple line) was recovered more accurately. (b) The separate analysis showed an improvement in the error in number of FC state transitions for G1 (dotted yellow line) when k was underestimated. (c) Unless k was correctly estimated, the combined analysis yielded an incorrect group difference in number of FC state transitions. (d) For the separate analysis, incorrect group differences were only found when k was overestimated or grossly misspecified for either group. (e) In terms of recovered space-time connectivity structure, the combined analysis performed better for G2 (solid purple line) than G1 (solid yellow line) when k was underestimated. (f) If k was correctly specified or only slightly misspecified for either group, the separate analyses had a similar error in recovered connectivity structure. Asterisks in (a-d) indicate a statistically significant ($p < 0.05$) difference from zero, according to a Wilcoxon signed-rank test, while asterisks in (e-f) indicate a statistically significant ($p < 0.05$) difference between the two groups, according to a two-sample Wilcoxon rank-sum test.

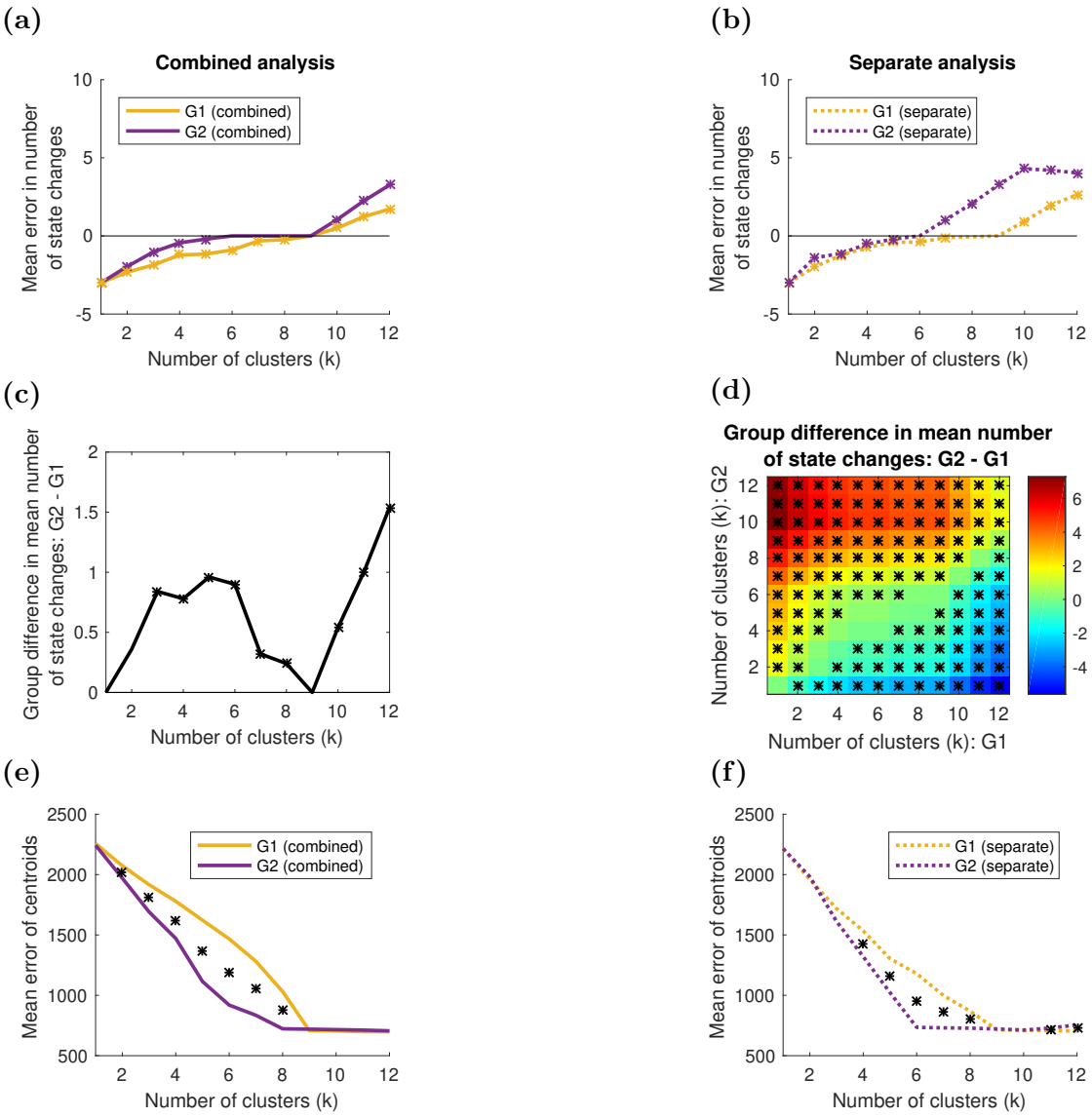


Figure 4.13: The results for the k-means simulation for window length $w = 90$. The true number of FC state transitions in this simulation is 3. Individuals in G1 could reach 9 FC states while individuals in G2 could reach only 6 of these 9 FC states. (a) When $k < 9$, the combined analyses underestimated the number of FC state transitions for G1 (solid yellow line), while the number of FC state transitions for G2 (solid purple line) was recovered more accurately. (b) The separate analysis showed an improvement in the error in number of FC state transitions for G1 (dotted yellow line) when k was underestimated. (c) Unless k was correctly estimated, the combined analysis yielded an incorrect group difference in number of FC state transitions. (d) For the separate analysis, incorrect group differences were only found when k was overestimated or grossly misspecified for either group. (e) In terms of recovered space-time connectivity structure, the combined analysis performed better for G2 (solid purple line) than G1 (solid yellow line) when k was underestimated. (f) If k was correctly specified or only slightly misspecified for either group, the separate analyses had a similar error in recovered connectivity structure. Asterisks in (a-d) indicate a statistically significant ($p < 0.05$) difference from zero, according to a Wilcoxon signed-rank test, while asterisks in (e-f) indicate a statistically significant ($p < 0.05$) difference between the two groups, according to a two-sample Wilcoxon rank-sum test.

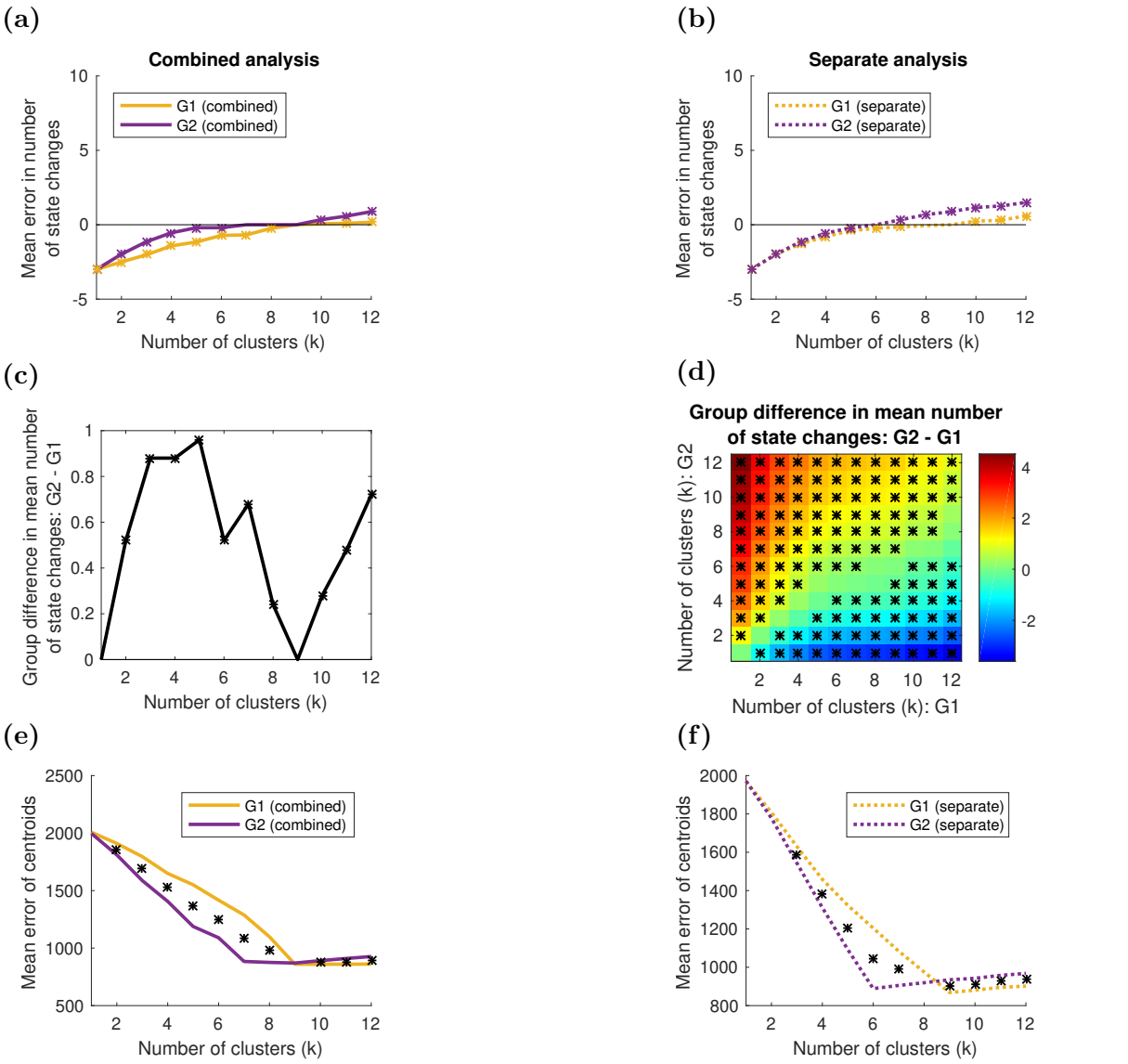


Figure 4.14: The results for the k-means simulation for window length $w = 120$. The true number of FC state transitions in this simulation is 3. Individuals in G1 could reach 9 FC states while individuals in G2 could reach only 6 of these 9 FC states. (a) When $k < 9$, the combined analyses underestimated the number of FC state transitions for G1 (solid yellow line), while the number of FC state transitions for G2 (solid purple line) was recovered more accurately. (b) The separate analysis showed an improvement in the error in number of FC state transitions for G1 (dotted yellow line) when k was underestimated. (c) Unless k was correctly estimated, the combined analysis yielded an incorrect group difference in number of FC state transitions. (d) For the separate analysis, incorrect group differences were only found when k was overestimated or grossly misspecified for either group. (e) In terms of recovered space-time connectivity structure, the combined analysis performed better for G2 (solid purple line) than G1 (solid yellow line) when k was underestimated. (f) If k was correctly specified or only slightly misspecified for either group, the separate analyses had a similar error in recovered connectivity structure. Asterisks in (a-d) indicate a statistically significant ($p < 0.05$) difference from zero, according to a Wilcoxon signed-rank test, while asterisks in (e-f) indicate a statistically significant ($p < 0.05$) difference between the two groups, according to a two-sample Wilcoxon rank-sum test.

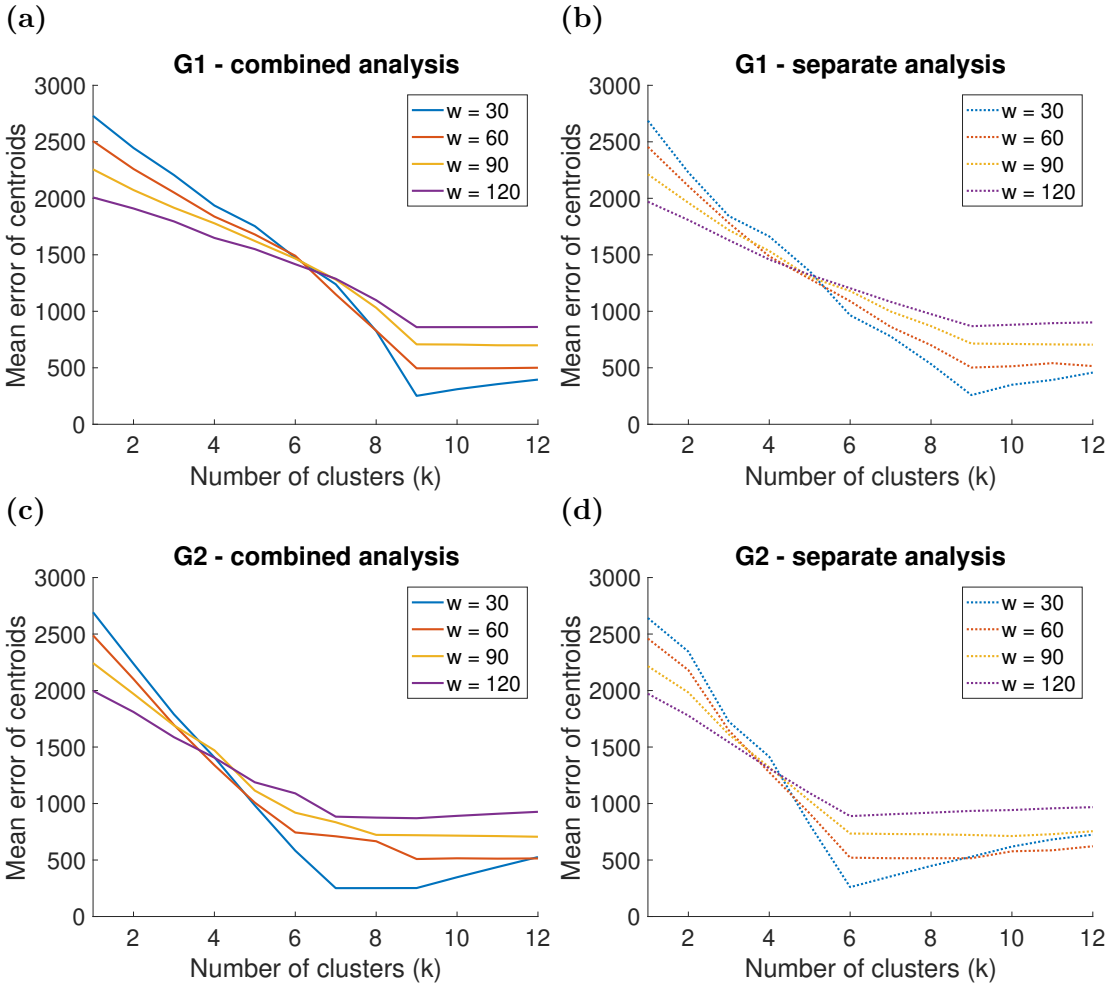


Figure 4.15: The effect of window length on mean (across subjects) error of centroids in the k-means simulation for (a) G1 in the combined analysis, (b) G1 in the separate analysis, (c) G2 in the combined analysis, and (d) G2 in the separate analysis. Individuals in G1 could reach 9 FC states while individuals in G2 could reach only 6 of these 9 FC states. Window length $w = 30$ is reported in the main text. These results show that when k is correctly estimated or only slightly misspecified, a longer window results in a worse estimation of the FC states, as evidenced by the increased mean centroid error.

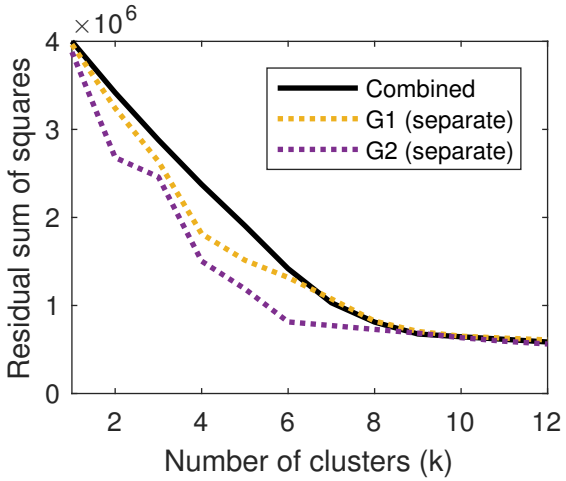


Figure 4.16: The RSS plots for the separate analysis of G2 has a clear elbow at $k = 6$ but there is no obvious elbow for the separate analysis of G1 or the combined analysis of both groups.

tivity structure is recovered. In this case, the combined analysis performed better for the young group than the old group when k was underestimated ($k < 9$), indicating that the recovered centroids were biased towards the 6 joint states. In contrast, if k was correctly specified or only slightly misspecified for either group, the separate analyses had a similar error in recovered connectivity structure. While these results suggest that the separate analyses did yield some improvement on the combined analysis, we caution that the problem of estimating k for both groups still needs to be addressed. As shown in Figure 4.11c and 4.11d, without an accurate estimation of k, one is likely to incorrectly infer the size of differences in dFC across groups, even for cases where there is no true difference. Figure 4.15 directly compares the mean centroid errors for different window lengths. While longer window lengths result in lower centroid errors for small k, shorter windows result in a better estimation of the FC states when k is close to the true number of FC states ($k = 9$ for G1, $k = 6$ for G2). In particular, the most accurate estimate of FC states occurs for shortest window length, $w = 30$, when k is correctly specified.

A common method for estimating the number of clusters, k, is the elbow plot shown in Figure 4.16. This demonstrates that it is not always straightforward to estimate k accurately: while we see a clear elbow for the separate analysis of G2 (purple dotted line), there is no obvious elbow for the separate analysis of G1 (yellow dotted line) or the combined analysis (solid black line).

4.4.6 Multilayer modularity

While the k-means method represents an approach which aggregates data across subjects in order to glean information about functional connectivity dynamics, other methods analyse fMRI data on a subject-by-subject basis. For example, the multilayer modularity approach [15] characterises the correlation matrices for a single subject obtained from a sliding-window analysis as a multi-layered network. Each region in each window is assigned a module label by maximising a modularity index which depends on two hyperparameters γ and ω (see Section 4.3.6 for details).

In this simulation, we again generated data for two groups of 50 individuals: those in G1 experienced one FC state transition while those in G2 experienced three FC state transitions. We applied the multilayer algorithm for all 100 individuals for $\gamma = 0.75, 1, \ldots, 2.5$ and $\omega = 0.25, 0.5, \ldots, 4$. One measure of dFC in this approach is the mean "flexibility" of each brain region (see Section 4.3.6). Thus, we now simulated a true group difference in the number of FC state transitions, and examined how accurately the FC states were recovered as a function of the method's hyperparameters. We also examined the error in the true vs estimated mean flexibility for each group.

Figure 4.17 illustrates the importance of parameter selection for the multilayer modularity approach. Figures 4.17a and 4.17c show that, in terms of the complete space-time connectivity dynamics, the optimal value for ω differed between the two groups. A lower value for ω generally resulted in more changes in module assignment across consecutive time windows. Since an individual in G1 experienced fewer FC state transitions, brain regions had fewer changes in module assignment across the course of the scan. Thus for G1, a higher value for ω was more effective in recovering the spatio-temporal connectivity structure.

We note that the optimal value for γ appears to be broadly the same for both groups. The parameter γ influences the resolution of the recovered network. A higher value for γ partitions the brain regions into more modules. Since ROIs were always partitioned into 5 modules for both groups, we would not expect the optimal value of γ to differ between the groups.

Figures 4.17b and 4.17d show that larger values of ω yield higher recovered mean flexibility for both G1 and G2. This effect, however, does not occur at the same rate for both groups. Figure 4.18 shows that different values of γ and ω result in different group differences in mean flexibility, to the extent that G1 is incorrectly found to be more flexible for some values of the hyperparameters. Note that the true difference in flexibility

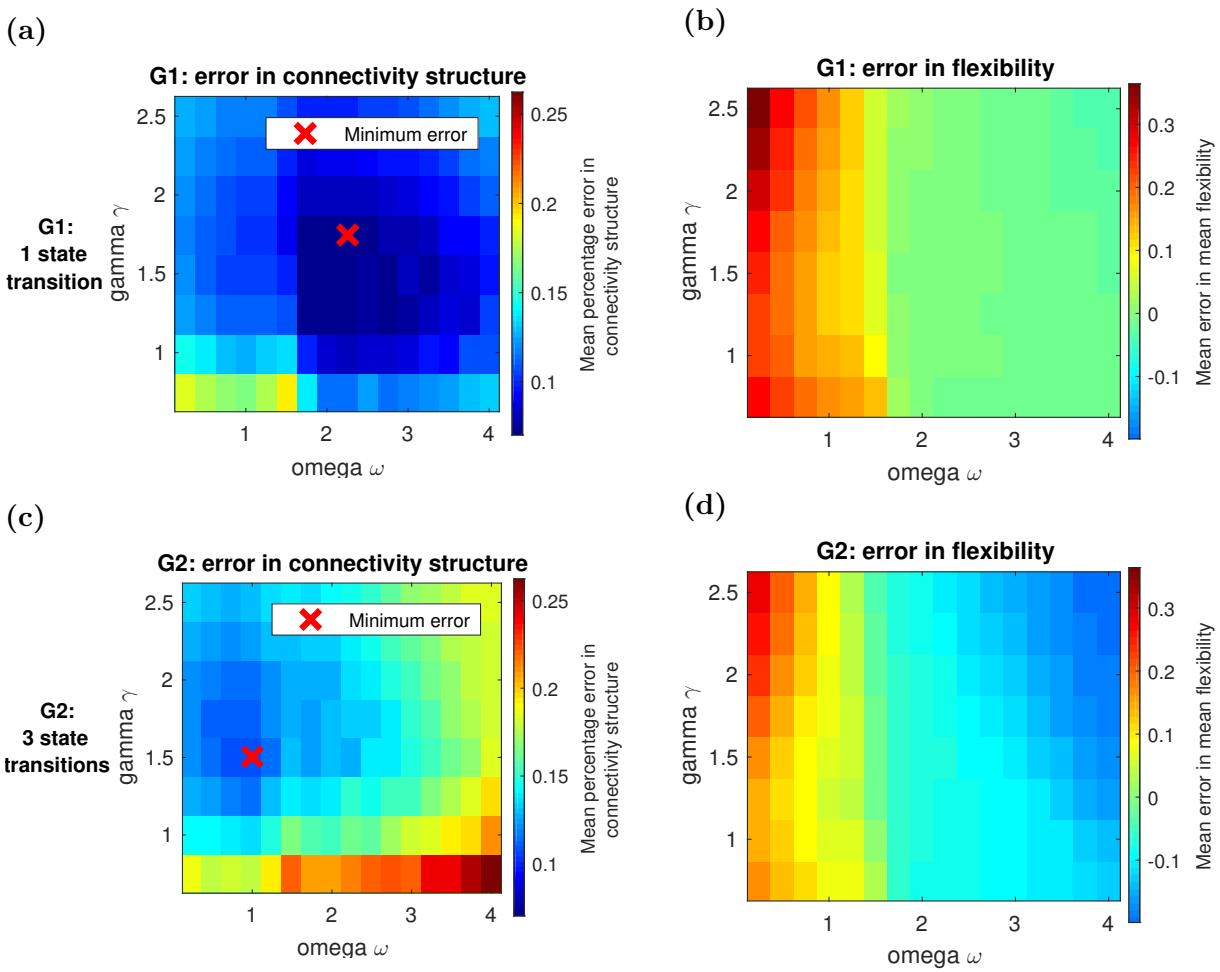


Figure 4.17: The effect of varying parameters γ and ω on performance of the multilayer modularity approach. Individuals in G1 experience 1 state transitions while individuals in G2 experienced 3 state transitions. Percentage error in connectivity structure is defined as the percentage of entries in the recovered incidence matrix that are equal to the corresponding entries of the true incidence matrix. A red cross indicates the pair of parameter values (γ, ω) which minimised the mean (across individuals) percentage error in connectivity structure. The flexibility of a region is defined as the number of times the region changes module assignment divided by the total possible number of module changes. (a) For G1, the optimal parameter values for recovering the connectivity structure were $\gamma = 1.25, \omega = 2$. (b) Increasing ω resulted in decreased recovered flexibility for G1. (c) For G2, the optimal parameter values for recovering the connectivity structure were $\gamma = 1.5, \omega = 1$. Thus, the optimal value for ω was markedly lower for G2 than for G1 while the optimal value for γ was slightly higher for G2 compared to G1. (d) Increasing ω also resulted in decreased recovered flexibility for G2, but at a greater rate than for G1.

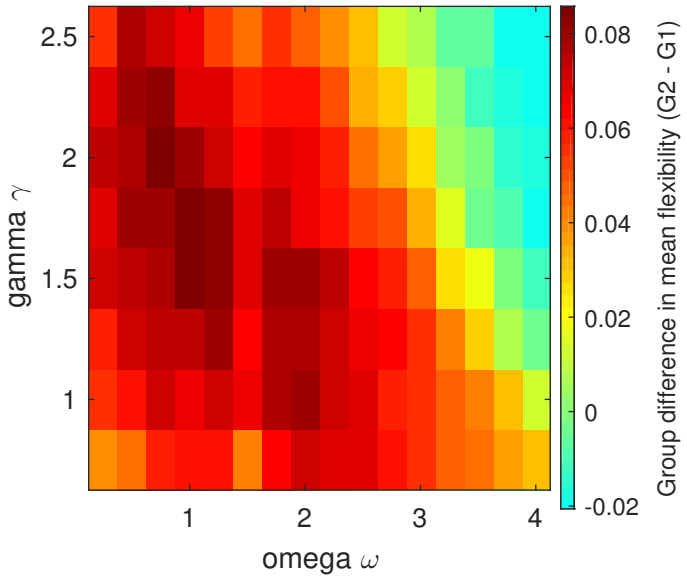


Figure 4.18: The effect of varying parameters γ and ω in the multilayer modularity approach on the recovered group difference in mean (across regions) flexibility. Individuals in G1 experience 1 state transitions while individuals in G2 experienced 3 state transitions. The flexibility of a region is defined as the number of times the region changes module assignment divided by the total possible number of module changes. Different values of γ and ω yield different group differences in mean flexibility. Note that the true difference here is 0.14, which is not attained for any values of the parameters.

was approximately 0.14 (see Section 4.3.6) and this was not captured for any values of γ and ω. This suggests that caution should be taken when computing group differences, especially when an assumption of homogeneity is made.

4.5 Discussion

We have illustrated some of the limitations of current dFC methods when dealing with heterogeneity. We used a generic simulation framework to isolate various sources of heterogeneity, and showed that observed connectivity dynamics may be due to factors other than true changes in connectivity.

To investigate the effects of individual differences in neural autocorrelation, HRF shape, connectivity strength and measurement noise, we used the SD of correlation values across sliding windows as our measure of dFC. We calculated this measure for three types of connectivity: static, positively connected; static, unconnected; and dynamically connected (positively connected to unconnected). Increased neural autocorrelation resulted

in higher dFC for statically connected regions but lower dFC for dynamically connected regions. A more temporally dispersed HRF produced higher dFC for all three connectivity types. In contrast, increased measurement noise yielded lower dFC across the three types of connectivity. Increased connectivity strength resulted in higher dFC for the dynamically connected regions but lower dFC for the positively statically connected regions. Together, these findings demonstrate that individual differences in dFC can be caused by various properties of the fMRI signal that are unrelated to the underlying neural connectivity dynamics.

We also demonstrated that common dFC methods may detect artifactual group differences in dynamic connectivity due to the assumptions that are made. For example, in a k-means analysis, it is often assumed that all individuals may attain the same set of FC states. If the hyperparameter k is incorrectly estimated, an incorrect group difference in the number of FC state transitions experienced may be detected if one group can attain more FC states than the other group. On the other hand, if the correct k is specified, the recovery of the number of FC state transitions is highly accurate. This suggests that it is the k-means clustering operation itself (rather than the correlation measure or windowing procedure) that leads to the observed biases when inhomogeneous groups are compared. In particular, the biases are likely due to the tendency of the k-means algorithm to detect the most prevalent states. We note that these issues could in principle affect any FC state-based method which assumes homogeneity in attainable FC states across individuals.

More generally, care should be taken with any method that requires the selection of hyperparameters by the user. In particular, we demonstrated that group differences in mean flexibility detected by a multilayer modularity approach were regulated by the choice of hyperparameters. While one would expect individual-based methods such as the multilayer modularity approach to be more robust to heterogeneity, spurious group differences can nonetheless be found if hyperparameters are assumed constant across individuals.

One could attempt to optimise the choice of hyperparameters using the data. For example, in a k-means analysis, the number of clusters, k, could be estimated by a variational Bayes approach [58]. For the multilayer modularity approach, one could use cross-validation across independent scans in an attempt to maximise stability of the recovered connectivity structure, though this assumes that dynamics are invariant across scans on the same individual. Alternatively the values of γ and ω could be chosen to yield connectivity structure that is most different from particular null models [14]. Hyperpa-

rameter optimisation could be investigated in the future work, though our point is that such optimisation should allow for heterogeneity across individuals.

We focused on a number of likely sources of heterogeneity in fMRI signals, using effects of age to illustrate some of our examples. This is based on recent evidence of group differences in signal autocorrelation [51, 10], HRF shape [69, 73, 3, 39], and non-neural physiological noise levels [53, 96]. Nevertheless, our findings apply in any situation where such heterogeneity may arise between individuals. Furthermore, there may be other sources of heterogeneity not investigated here that could have spurious effects on observed dFC. For example, we only considered variability in 2 out of the 7 HRF parameters; it is plausible that the remaining parameters also have an effect on estimated dFC. Similarly, we assumed that brain regions partition into 5 modules in each FC state, whereas it is conceivable that this could differ among individuals.

We note that certain aspects of the simulation framework represent simplifications of the physical and physiological processes involved in fMRI neuroimaging. For example, the addition of white measurement noise is not realistic and noise related to head motion or vascular effects may have more regionally specific effects on connectivity estimates [112]. Therefore the noise simulation should be interpreted as a cautionary result, and not as an illustration of effects of real noise sources in fMRI. We also assumed that the HRF is the same for all regions, although the shape of the HRF has been shown to vary from region to region (e.g. [123]). These assumptions, however, allowed us to isolate the impact of unaccounted heterogeneity. In particular, our current simulation framework had the distinct benefit of separating the underlying dFC structure from sources of heterogeneity such as neural autocorrelation, noise and HRF shape. We were thus able to manipulate dFC and these other sources of variation independently and show how observed dFC is affected. Further, we demonstrate that the effects of unrelated sources of variation persist across a range of parameter values, suggesting that these effects of dFC are not specific to this simulated dataset, but are in fact a more general phenomenon.

We do not address what drives these changes in functional connectivity. Recent observations suggest that dFC can be explained in terms of sparse brief events [6, 79, 93, 142]. Our simulation framework is based on work a previous study [7] that attempted to find periods of recurrent patterns of functional connectivity, or FC states, across time and individuals. It may be that these periods are longer than the underlying neural processes due to the temporal limitations of fMRI. Many FC state-based studies work on the basis of certain assumptions about FC states regarding, for example, their discreteness and

typical duration. Generalisations of these approaches exist: FC states can be separated across distinct frequency profiles [166], or functional connectivity can be characterised using a meta-state analysis, which allows different FC states to occur simultaneously [99] . However, little is actually known about the nature of FC states and future work is required to better understand how these neural processes drive observed FC states.

To illustrate the effects of heterogeneity in the number of attainable states and the number of state changes, we simulated binary differences between two groups of individuals. This represents a simplification since, in reality, it is likely that differences between individuals fall on a continuous spectrum and so we caution against dichotomising between groups. It should also be emphasised that while here we have isolated the impacts of different sources of heterogeneity, in reality they may appear in combination. Future work could investigate how different types of heterogeneity interact, or even counteract, to produce differences in observed dFC.

Our work does not specify precisely how the different aspects of an analysis pipeline affect artifactual differences in observed dFC. Many of the sources of heterogeneity presented in this paper are also likely to affect estimates of static functional connectivity, as measured by correlation. For example, static FC decreases with higher white noise and is modulated by HRF shape [91]. Therefore, connectivity estimates for each window in a dFC analysis will be affected by unmodelled heterogeneity and combine (possibly in a non-linear fashion) to influence the final dFC estimates. Recent work has aimed to disentangle changes in FC from changes in signal level [40], but an important next step is to determine the precise manner in which observed dFC is affected by the different parts of the dFC analysis pipeline, with the aim to develop methods that resolve the issues raised by our findings.

Although we have chosen to illustrate the above points with only a few methods, the issues should in general extend to other approaches. For example, we used Fisher-transformed Pearson correlation as our basic measure of functional connectivity because this is currently the most commonly used metric. Alternative connectivity measures, such as coherence or multiplication of temporal derivatives [129], may be less susceptible to certain types of unaccounted heterogeneity: for example, it has been shown that coherence is robust against variability in the shape of the HRF between regions [12]. Nonetheless, the issues of hyperparameter selection in dFC methods still need to be addressed regardless of the connectivity metric used.

4.6 Future work

The main contribution of this chapter was to develop a simulation framework for fMRI time series driven by a dynamic connectivity structure. We used this framework to isolate various sources of heterogeneity that are likely to be present in real data, and showed that observed changes in connectivity may be due to factors other than true changes in connectivity. We identify two potential avenues for future work into the development of dFC methods. Firstly, more work is needed to understand how different sources of heterogeneity can affect connectivity metrics. Our simulation framework is simple, serving to highlight that heterogeneity does have some effect on dFC measures, and may not be appropriate for determining the precise nature of those effects. Secondly, new methods are needed that correct or account for these various sources of heterogeneity in order to identify true differences in dFC between individuals or groups of individuals.

Chapter 5

Modelling dynamic functional connectivity with Bayesian exponential random graph models

Summary

As the importance of understanding dynamic functional connectivity becomes more apparent, there is a need for methods that are able to identify meaningful differences between groups. Just as static functional connectivity can naturally be modelled as a (static) network, dynamic functional connectivity can be characterised as a temporal network. In this Chapter, we extend the multilevel framework presented in Chapters 2 and 3 to the case of multiple networks per individual. Our motivation again stems from the Cam-CAN project. Using resting-state magnetoencephalography data from this study, we illustrate how our method can be used to detect differences in the variability of functional connectivity between a group of young individuals and a group of old individuals.

5.1 Background

The majority of dFC methods can be seen as extensions of a basic sliding-window analysis [113]. To recap, a sliding-window analysis proceeds by splitting the regional time series into (possibly overlapping) windows. Functional connectivity is then computed on a window-by-window basis. Typically, the connectivity pattern is described in terms of pairwise relations (e.g. correlation) between ROI time series, resulting in a set of connectivity matrices indexed by time for each subject. Methods to analyse the resulting connectivity matrices vary greatly; we now focus on network-based methods such as the multilayer modularity approach discussed in the previous chapter [15, 16].

Notation

Before reviewing existing network approaches for characterising dynamic functional connectivity, we first recap and introduce some notation. Denote J to be the number of groups, and the number of individuals in group j as n_j (allowing for unbalanced groups, $n_1 \neq n_2$ and assuming that the group memberships are known). We assume that the set $\mathcal{V}$ of nodes is fixed and common among all networks in all groups, and let $V = |\mathcal{V}|$ be the number of nodes. We will work directly with the adjacency matrices associated with each individual's network. Let $\mathbf{Y}^{(t,i,j)}$ be the (random) adjacency matrix associated with individual i in group j at time t, $1 \leq t \leq T_{ij}$, and denote $\mathbf{y}^{(t,i,j)}$ to be an instantiation of $\mathbf{Y}^{(t,i,j)}$. Thus, $\mathbf{Y}^{(t,i,j)} \in \{0,1\}^{V \times V}$ and $Y_{kl}^{(t,i,j)} = 1$ indicates a connection between regions k and l. We allow for a different number of networks per individual, $T_{ij} \neq T_{i'j'}$ for $(i,j) \neq (i',j')$. We assume that the range of the $\mathbf{Y}^{(t,i,j)}$, i.e. the set of all possible outcomes, is common across all networks and denote this $\mathcal{Y}$. Let $s(\mathbf{Y}^{(t,i,j)})$ denote a vector of p summary statistics of $\mathbf{Y}^{(t,i,j)}$, such that each component is a function $s_i : \mathcal{Y} \to \mathbb{R}$. Further, we denote $\boldsymbol{\theta}^{(I)} = (\theta^{(i,j)})_{j=1,\ldots,J, i=1,\ldots n_j}$ to be a vector of all individual-level parameters, $\boldsymbol{\theta}^{(T)} = (\theta^{(t,i,j)})_{t=1,\ldots,T_{ij}, j=1,\ldots,J, i=1,\ldots n_j}$ to be a vector of all network-level parameters, and $\boldsymbol{\theta}^{(T,i,j)} = (\theta^{(t,i,j)})_{t=1,\ldots,T_{ij}}$ to be a vector of all network-level parameters for individual i in group j.

Network approaches for dynamic functional connectivity

The simplest network approaches for characterising dFC proceed by computing graph metrics separately for each window. Connectivity dynamics can then be assessed by

studying the resulting graph metric time series [143, 169]. A similar family of methods uses tools from temporal network theory [72]. These methods treat the entire set of window-by-window networks as a single temporal network. One can then compute a variety of temporal graph metrics to describe the whole dynamic connectivity structure [135]. Examples previously used in neuroimaging include multilayer modularity [15, 16] and temporal efficiency [147]. These metric-based approaches suffer from the same drawback discussed in Chapter 3 in the context of static networks. In particular, fluctuations in these graph metrics may simply reflect variations in the network density. Given the results from the previous Chapter, this indicates that group-level differences in these metrics may arise due to differences in non-dynamic features of the data such as noise levels.

Another class of methods uses temporal extensions of statistical network models. Some examples include a dynamic stochastic block model used to describe changes in community structure in a drug-induced analgesia fMRI experiment [119], and a separable temporal exponential random graph model [85] applied to analyse functional network dynamics in a recovering coma patient [36]. In both these cases, however, the models are applied to only a single network obtained either from a single subject or an average over multiple subjects. In order to capture variability in dFC between subjects, it is necessary to model each (temporal) network separately.

Motivated by this, we show how the multilevel framework presented in Chapters 2 and 3 can be used to jointly model dynamic functional connectivity networks. We illustrate our method using resting-state MEG data from the Cam-CAN study [127], described in Section 1.3. In Section 5.3, we extend the framework to deal with multiple networks per individual. Section 5.4 presents the results from applying our framework to the Cam-CAN MEG data, and Section 5.5 discusses the limitations of our approach.

5.2 Data

To illustrate our method, we use resting-state MEG data from the Cam-CAN study. The decision to shift our attention from fMRI to MEG data is motivated by our findings from the previous chapter: MEG is not susceptible to effects from the haemodynamic response function.

The data were collected as part of Phase II of the CamCAN project (www.cam-can.org; [127]). The MEG data were acquired on a 306-channel Vectorview system (Elekta Neuromag, Helsinki), situated in a light, magnetically-shielded room at the MRC Cognition &

Brain Sciences Unit. This system contains two orthogonal planar gradiometers and one magnetometer at each of 102 positions. Data were acquired continuously at 1 kHz. Head-Position Indicator (HPI) coils were used to monitor head position. Electrooculograms were recorded using paired EOG electrodes in a bipolar montage. The three-dimensional locations of the HPI coils, over 100 'head points' across the scalp, and three anatomical fiducials (the nasion and left and right pre-auricular points) were recorded using a 3D Fastrak digitizer (Polhemus Inc., Colchester, VA). The current data are from 8 min and 40 s of resting with eyes-closed (with same instructions as for the fMRI data in Chapter 3).

A structural MRI image was used for constructing the forward model that maps from unknown cortical electrical activity to the magnetic fields/gradients recorded by the MEG system modelling the head. This structural image was acquired on a 3T Siemens TIM Trio at the MRC Cognition & Brain Sciences Unit using a 1mm3 isotropic (MPRAGE) sequence, as described in Chapter 3.

To assess how the dynamic functional connectivity differs between young individuals and older individuals, we again focused on the youngest (18-27, 46 subjects) and oldest (78-88, 89 subjects) decades. These subjects are a subset of those in Chapter 2. A total of 12 subjects were removed owing to either problems in preprocessing or coregistration of their MEG data with their MRI data (see below). Section 5.2.1 outlines the preprocessing steps undertaken to prepare the data for analysis. Section 5.2.2 describes the sliding-window analysis used to construct temporal networks from the time series data for each individual.

5.2.1 Preprocessing

The MEG data come from local Release0002 of the CamCAN MEG pipelines. These pipelines used r5.4 of the Automatic Analysis (AA) pipeline system (Cusack et al., 2014; for overview of pipelines, see Taylor et al., 2017) in r2015 of MATLAB (The MathWorks). Preprocessing started with application of the Elekta MaxFilter (v2.2.12) software that implements the signal space separation (SSS) approach of Taulu et al (2005). The origin of the SSS expansion was determined by fitting a sphere to all digitized head points, except those around the nose (i.e., with $y > 0$ and $z < 0$, where y and z refer to anterior-posterior and inferior-superior axes of the head, with the origin defined as the intersection of lines between fiducials and nasion). Mains noise at 50Hz and its harmonics were attenuated. Bad channels were detected and reconstructed automatically (with standard deviation criterion of 7). The temporal extension of SSS was used with 10s windows and an inner-

outer sphere correlation criterion of 0.98. The data were corrected for head motion every 1s. All the raw and maxfiltered data for this dataset, together with more acquisition details, are available on: http://www.cam-can.org/index.php?content=dataset.

The resulting maxfiltered files were then converted into Matlab using r7219 of the SPM12 software (http://www.fil.ion.ucl.ac.uk/spm). The MEG data were coregistered to the structural MRI using the three digitised fiducials, after rigid-body registration of the structural MRI image and fiducials to MNI space. A forward model was created using a single-shell model. The data were then downsampled to 250Hz for computational efficiency and bandpass filtered at 1-100 Hz (using 5th order Butterworth filter).

The OSL software (https://ohba-analysis.github.io/osl-docs/) was then used for the final steps. The data were transformed from sensor space to source space (3559 voxels) using a linearly constrained minimum variance beamformer [162]. This particular beamforming method uses Bayesian principal component analysis (PCA) to regularize the data covariance matrix estimation. Next, the beamformed data were split into the same 38 ROI parcellation as used in Chapter 3. A representative time series for each ROI was constructed by taking the coefficients of the first principal component. These time series were reweighted to account for the number and the variance of the voxels in each ROI. Finally, each ROI time series was bandpass filtered into the frequency bands of interest: θ (4Hz to 7Hz), α (8Hz to 13Hz), β (13Hz to 30Hz), low γ (30Hz to 50Hz), and high γ (50Hz to 100Hz).

5.2.2 Sliding-window analysis

The following analysis was performed separately for each of the frequency bands. To assess how amplitude correlation varied over the course of a scan, we split each time series into non-overlapping windows of width $w = 60s$. This typically resulted in $T = 9$ windows for each individual.

Before computing amplitude correlation in each window, we performed a signal leakage reduction step. Signal leakage occurs as a result of the source reconstruction process. Since source activity for several thousand voxels is being estimated from only a few hundred sensors, the reconstruction of true point sources will be spread over several voxels, resulting in spurious correlations between ROIs. Signal leakage reduction was performed via multivariate orthogonalisation of the ROI time series in each window, thus removing zero-lag correlations between ROIs [30]. Following signal leakage reduction, we computed the amplitude correlation between each pair of ROIs in each window t for each indi-

vidual i in group j, yielding a $N \times N$ connectivity matrix $C^{(t,i,j)}$ where $N = 38$ is the number of ROIs (the same ROIs used for the fMRI data in Chapter 3). Recall that amplitude correlation measures the Pearson correlation between the logarithm of the *amplitude envelopes*. While the signal leakage reduction step removes zero-lag correlations, the remaining correlations between the amplitude envelopes are thought to indicate genuine biological dependencies.

To convert these connectivity matrices into networks, we considered three thresholding procedures. The first approach, absolute thresholding, uses a constant threshold across all subjects and all windows. This approach, however, resulted in empty networks for some windows and complete networks for others, due to the variability in mean connectivity across windows and subjects. The second approach, proportional thresholding, uses a different threshold for each window and subject. Given the interest in changes functional connectivity, this approach is less appropriate as it maintains the same density across all the networks.

The third approach (our preferred option), individual-based thresholding, ensures the same *average* network density across windows for each subject but allows density to change between windows. Specifically, we applied a threshold $r^{(i,j)}$ to each correlation matrix to produce an $N \times N$ adjacency matrix A with entries:

$$A_{kl}^{(t,i,j)} = \begin{cases} 1 & \text{if } C_{kl}^{(t,i,j)} \geq r^{(i,j)} \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

The thresholds $r^{(i,j)}$ are chosen such that average network density for an individual is the same across all individuals. As there is no objective way to choose this average network density, we considered two values: 10% and 20%. Note that this procedure results in networks $\mathbf{Y}^{t,i,j}$ for $t = 1, \ldots, T_{ij}, i = 1, \ldots, N_j, j = 1, \ldots, G$ corresponding to the adjacency matrices $\mathbf{A}^{(t,i,j)}$.

5.3 Methods

In Chapter 2, we proposed a multilevel framework to model a population of networks. To briefly recap, this framework models each network $\mathbf{Y}^{(i)}$ as an exponential random graph with individual-level parameter $\theta^{(i)}$. Next, we assume that the individual-level parameters are draw from a group-level distribution $\theta^{(i)} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. In Section 3.3.1, we extended

this framework to deal with group structure within the population. We now describe a further extension to the framework to handle several networks per individual, as generated by the sliding-window analysis above.

General framework

The extension amounts to adding another level to the hierarchy of the model. Each network $\mathbf{Y}^{(t,i,j)}$ is modelled as an ERGM with respective parameter $\theta^{(t,i,j)}$. In turn, we assume each ERGM parameter is drawn from an individual-level distribution:

$$\theta^{(t,i,j)} \sim \mathcal{N}(\theta^{(i,j)}, \mathbf{\Sigma}^{(i,j)}), \quad t = 1, \ldots, T_{ij}, \quad i = 1, \ldots, N_j, \quad j = 1, \ldots, G \tag{5.2}$$

Note that the $\mathbf{\Sigma}^{(i,j)}$ covariance parameter characterises the variability of the connectivity structure for individual i in group j.

The remainder of the model is as before. To recap, we assume the individual-level means are drawn from a group-level normal distribution $\theta^{(i,j)} \sim \mathcal{N}(\mu^{(j)}, \mathbf{\Sigma}^{(j)}), \quad i = 1, \ldots, N_j, \quad j = 1, \ldots, G$. Next, we assume the group-level means are then drawn from a population-level normal distribution $\mu^{(j)} \sim \mathcal{N}(\mu^P, \mathbf{\Sigma}^P), \quad j = 1, \ldots, G$. The full model is depicted as a Bayesian network in Figure 5.1.

Adjusted framework

Since there are only around $T_i = 10$ networks per individual, the individual-level covariance parameter $\mathbf{\Sigma}^{(i,j)}$ is likely to be poorly estimated. To bypass this issue, we make the simplifying assumption that the individual-level covariance is the same across individuals within a group. Specifically, we assume

$$\theta^{(t,i,j)} \sim \mathcal{N}(\theta^{(i,j)}, \mathbf{\Sigma}^{(j)}), \quad t = 1, \ldots, T_{ij}, \quad i = 1, \ldots, N_j, \quad j = 1, \ldots, G. \tag{5.3}$$

Thus, $\mathbf{\Sigma}^{(j)}$ characterises the variability in functional connectivity for the whole of group j. One way to quantify the variability is through the matrix determinant $|\mathbf{\Sigma}^{(j)}|$. Via the posterior samples of $|\mathbf{\Sigma}^{(j)}|$, we are able to compare dFC across groups. We again make the assumption that the variability of the individual-level means are the same across the groups. In other words, we assume $\theta^{(i,j)} \sim \mathcal{N}(\mu^{(j)}, \mathbf{\Sigma}_\theta)$. The full model for this adjusted framework is depicted in Figure 5.2.

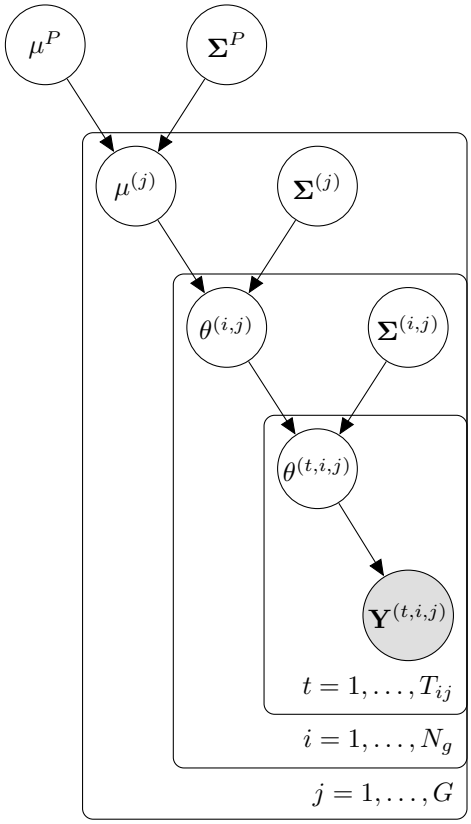


Figure 5.1: Diagrammatic representation of the general multilevel model for networks generated from a sliding-window analysis. The network $\mathbf{Y}^{(t,i,j)}$ constructed from the t^{th} window of the i^{th} individual in the j^{th} group is modelled as an ERGM with window-level parameter $\theta^{(t,i,j)}$. This is assumed to be drawn from an individual-level multivariate normal distribution with mean $\theta^{(i,j)}$ and covariance $\mathbf{\Sigma}^{(i,j)}$. In turn the individual-level mean is drawn from a group-level multivariate normal distribution with mean $\mu^{(j)}$ and covariance $\mathbf{\Sigma}^{(j)}$. Finally, the group-level mean is assumed to be drawn from a population-level multivariate normal distribution with mean μ^P and covariance $\mathbf{\Sigma}^P$.

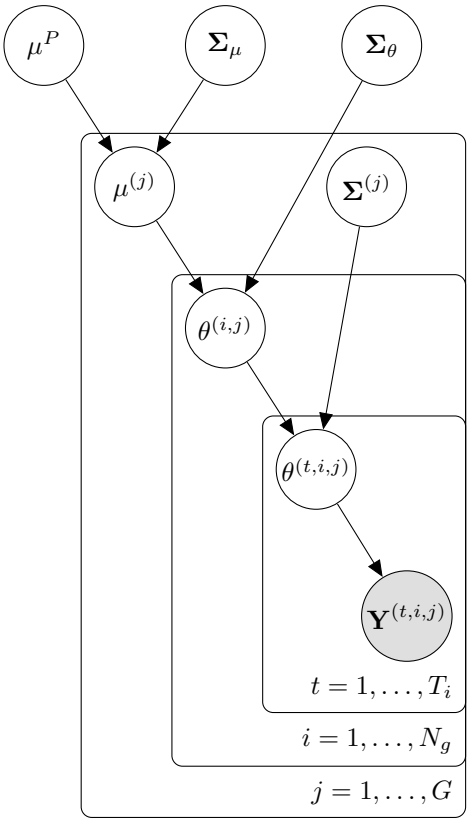


Figure 5.2: Diagrammatic representation of the adjusted multilevel model for networks generated from a sliding-window analysis. As with the general framework, the network $\mathbf{Y}^{(t,i,j)}$ constructed from the t^{th} window of the i^{th} individual in the j^{th} group is modelled as an ERGM with window-level parameter $\theta^{(t,i,j)}$. This is also assumed to be drawn from a multivariate normal distribution with an individual-level mean $\theta^{(i,j)}$ but a *group-level* covariance $\boldsymbol{\Sigma}^{(j)}$. This contrasts with the individual-level covariance in the general framework. The individual-level mean is then assumed to be drawn from a multivariate normal distribution with a group-level mean $\mu^{(j)}$ and a population-level covariance $\boldsymbol{\Sigma}_\theta$. Finally, the group-level mean is again drawn from a population-level multivariate normal distribution with mean μ^P and covariance $\boldsymbol{\Sigma}_\mu$.

Prior specification

To complete the model specification, we need to place prior distributions on the hyperparameters $(\mu^P, \mathbf{\Sigma}_\mu, \mathbf{\Sigma}_\theta, \mathbf{\Sigma}^{(1)}, \mathbf{\Sigma}^{(2)})$. As before, we used informative priors based on independent Bayesian ERGM fits across all $n = 625$ individuals for which resting-state MEG was measured. These individual fits resulted in $K = 12000$ posterior samples $\{\theta_k^{(t,i)}\}_{k=1}^K$ for each individual $i \in \{1, \ldots, 625\}$ and each of their respective time windows $t \in \{1, \ldots, t_i\}$. From these posterior samples, we calculated window-level posterior mean estimates $\bar{\theta}^{(t,i)} = \frac{1}{K}\sum_{k=1}^K \theta_k^{(t,i)}$, individual-level posterior mean estimates $\bar{\theta}^{(i)} = \frac{1}{KT_i}\sum_{k=1}^K \sum_{t=1}^{T_i} \theta_k^{(t,i)}$, and a population-level posterior mean estimate $\bar{\mu}^P = \frac{1}{n}\sum_{i=1}^n \frac{1}{KT_i}\sum_{k=1}^K \sum_{t=1}^{T_i} \theta_k^{(t,i)}$. The prior for the population-level mean parameter was set as:

$$\mu^P \sim \mathcal{N}(\bar{\mu}^P, 5I) \tag{5.4}$$

We placed the same priors on the population-level covariance parameters:

$$\mathbf{\Sigma}_\mu, \mathbf{\Sigma}_\theta \sim \mathcal{W}^{-1}(\hat{\mathbf{\Sigma}}^P, p+1) \tag{5.5}$$

where $p = 3$ is the number of summary statistics in the ERGM, and

$$\hat{\mathbf{\Sigma}}^P = \frac{1}{n-1}\sum_{i=1}^n (\bar{\theta}^{(i)} - \bar{\mu}^P)^T(\bar{\theta}^{(i)} - \bar{\mu}^P).$$

Finally, for the group-level covariance parameters, we placed the prior:

$$\mathbf{\Sigma}^j \sim \mathcal{W}^{-1}(\hat{\Sigma}^G, p+1) \tag{5.6}$$

where $\hat{\Sigma}^G = \frac{1}{n}\sum_{i=1}^n \frac{1}{T_i}\sum_{t=1}^{T_i}(\bar{\theta}^{(t,i)} - \bar{\theta}^{(i)})^T(\bar{\theta}^{(t,i)} - \bar{\theta}^{(i)})$.

Model fitting

To generate samples from the joint posterior, we further extend Algorithm 3.1 to deal with the case of multiple networks per individual. The updates for the population-level parameters $(\mu^P, \mathbf{\Sigma}_\theta, \mathbf{\Sigma}_\mu)$ and the group-level parameters $\mu^{(1)}, \mu^{(2)}, \mathbf{\Sigma}^{(1)}, \mathbf{\Sigma}^{(2)}$ are performed by sampling directly from their conditional posteriors, which is made possible due to the choice of conditionally conjugate prior distributions. Following this, the individual-level parameters $\theta^{(i,j)}$ and the window-level parameters are updated via an ancillarity-

sufficiency interweaving strategy. The full MCMC procedure is given in Algorithm 5.1.

Initialisation

To reduce the burn-in period (see Section 2.5.3), we initialised the algorithm using estimates of various mean parameters based on the individual Bayesian ERGM fits. Writing $\{\theta_k^{(t,i,j)}\}_{k=1}^K$ for the posterior samples of the fit to the t^{th} network of the i^{th} individual in the j^{th} group, we set

$$\begin{aligned}
\theta_0^{(t,i,j)} &= \frac{1}{K}\sum_{k=1}^{K}\theta_k^{(t,i,j)}, \quad j=1,2, \quad i=1,\ldots,n_j, \quad t=1,\ldots,T_i \\
\theta_0^{(i,j)} &= \frac{1}{T_i}\sum_{t=1}^{T_i}\theta_0^{(t,i,j)}, \quad j=1,2, \quad ,i=1,\ldots,n_j \\
\mu_0^{(j)} &= \frac{1}{n_j}\sum_{i=1}^{n_j}\theta_0^{(i,j)}, \quad j=1,2 \\
\mu_0^{P} &= \frac{1}{2}\sum_{j=1}^{2}\mu_0^{(j)}
\end{aligned} \tag{5.7}$$

and $\tilde{\theta}_0^{(t,i,j)} = \theta_0^{(t,i,j)} - \theta_0^{(i,j)}$ for $j=1,2, \quad i=1,\ldots,n_j, \quad t=1,\ldots,T_i$.

Proposals

To improve mixing of the MCMC, we used proposals based on the same individual Bayesian ERGM fits (see Section 2.5.2). For the window-level parameter updates, we drew proposals

$$\theta' \sim \mathcal{N}(0, \tau_T \hat{\mathbf{\Sigma}}_T) \tag{5.8}$$

where $\hat{\mathbf{\Sigma}}_T = \frac{1}{n_1+n_2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\frac{1}{T_i}\sum_{t=1}^{T_i}\frac{1}{K-1}\sum_{k=1}^{K}(\theta_k^{(t,i,j)} - \bar{\theta}^{(i,j)})(\theta_k^{(t,i,j)} - \bar{\theta}^{(i,j)})^T$ and τ_T is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4. For the individual-level parameter updates, we draw proposals centered at the current value θ:

$$\theta' \sim \mathcal{N}(\theta, \tau_I \hat{\mathbf{\Sigma}}_I) \tag{5.9}$$

where $\hat{\mathbf{\Sigma}}_I = \frac{1}{n_T-1}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\sum_{t=1}^{T_i}(\bar{\theta}^{(i,j)} - \bar{\theta}^T)(\bar{\theta}^{(i,j)} - \bar{\theta}^T)^T$ and τ_I is a tuning parameter again chosen to yield acceptance rates between 0.1 and 0.4. Here, $n_T = \sum_{j=1}^{2}\sum_{i=1}^{n_j}\sum_{t=1}^{T_i} 1$ is the total number of networks across the two groups, and $\bar{\theta}^T = \frac{1}{n_T}\sum_{j,i,t}\bar{\theta}^{(t,i,j)}$ is the

Algorithm 5.1 The ancillarity-sufficiency interweaving strategy (ASIS) algorithm for a two-group multilevel Bayesian ERGM with multiple networks per individual.

Require: number of MCMC iterations K, initial values $(\mu_0^P, \boldsymbol{\mu}_0, \boldsymbol{\theta}_0^{(I)}, \tilde{\boldsymbol{\theta}}_0^{(T)})$

for $k = 1, \ldots, N$ **do**

- draw $\boldsymbol{\Sigma}_{\theta,k} \sim \pi(\cdot|\tilde{\boldsymbol{\theta}}_{k-1}^{(I)}, \boldsymbol{\mu}_{k-1})$, $\quad \boldsymbol{\Sigma}_{\mu,k} \sim \pi(\cdot|\boldsymbol{\mu}_{k-1}, \mu_{k-1}^P)$, $\quad \mu_k^P \sim \pi(\cdot|\boldsymbol{\Sigma}_{\mu,k}, \boldsymbol{\mu}_{k-1})$

for $k = 1, 2$ **do**

- draw $\mu_k^{(j)} \sim \pi(\cdot|\mu_k^P, \boldsymbol{\Sigma}_{\mu,k}, \boldsymbol{\theta}_{k-1}^{(I)})$, $\quad \boldsymbol{\Sigma}_k^{(j)} \sim \pi(\cdot|\tilde{\boldsymbol{\theta}}^{(T,I,j)})$

end for

for $j = 1, 2, \quad i = 1, \ldots, n_j, \quad t = 1, \ldots, T_i$ **do**

- set $\theta_{k-1}^{(t,i,j)} = \theta_{k-1}^{(i,j)} + \tilde{\theta}_{k-1}^{(t,i,j)}$
- draw $\theta' \sim h_T(\cdot|\theta_{k-1}^{(t,i,j)})$
- draw $\mathbf{y}' \sim \pi(\cdot|\theta')$
- set $\theta_{k-0.5}^{(i,j)} = \theta'$ with probability

$$\min\left(1, AR(\theta', \theta_{k-1}^{(t,i,j)}; \theta_{k-1}^{(i,j)}, \boldsymbol{\Sigma}_{\theta,k}, \mathbf{y}^{(t,i,j)}, \mathbf{y}')\right)$$ ▷ See Eq. (2.31)

- else, set $\theta_{k-0.5}^{(t,i,j)} = \theta_{k-1}^{(t,i,j)}$

end for

for $j = 1, 2, \quad i = 1 \ldots, n_j$ **do**

- draw $\theta_{k-0.5}^{(i,j)} \sim \pi(\cdot|\mu_k^{(j)}, \boldsymbol{\theta}_{k-0.5}^{(T,i,j)}, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{\Sigma}_k^{(j)})$

end for

for $j = 1, 2, \quad i = 1, \ldots, n_j, \quad t = 1, \ldots, T_i$ **do**

- set $\tilde{\theta}_k^{(t,i,j)} = \theta_{k-0.5}^{(t,i,j)} - \theta_{k-0.5}^{(i,j)}$

end for

for $j = 1, 2, \quad i = 1, \ldots, n_j$ **do**

- draw $\theta' \sim h_I(\cdot|\theta_{k-0.5}^{(i,j)})$
- draw $\mathbf{y}^{(t)'} \sim \pi(\cdot|\theta' + \tilde{\theta}_k^{(t,i,j)})$ for $t = 1, \ldots, T_i$
- set $\theta_k^{(i,j)} = \theta'$ with probability

$$\min\left(1, AR(\theta', \theta_{k-0.5}^{(i,j)}; \tilde{\boldsymbol{\theta}}_k^{(T,i,j)}, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{y}^{(T,i,j)}, \boldsymbol{y}')\right)$$ ▷ See Eq. (2.33)

- else, set $\theta_k^{(i,j)} = \theta_{k-0.5}^{(i,j)}$

end for

end for

average across the respective posterior means.

5.4 Results

We performed our analysis across five frequency bands (θ, α, β, low γ, high γ) and two threshold values (10% and 20%), resulting in ten separate sets of results. For each frequency band and each threshold, we ran the algorithm for $K = 25000$ iterations. We discarded the first 5000 of these iterations following inspection of the trace plots, which indicated that convergence for each of the ten runs had occurred after this burn-in period.

Prior predictive checks

To assess the suitability of the prior distributions, we performed two prior predictive checks. Firstly, to check the mean of the ERGM parameters, we simulated 2000 networks from the population-level prior distribution and compared the distribution of summary statistics of this simulated population with that of the observed population of networks (Figures 5.3 and 5.4). While the distribution of the observed summary statistics are strongly peaked for both the young group (red) and the old group (blue), the distribution of the simulated statistics (grey) is much flatter, indicating that the prior distribution is not overinformative.

Despite a (unimodal) multivariate normal prior on the ERGM parameters, the prior predictive distribution of the resulting summary statistics appear to be bimodal (GWNSP statistic) if not trimodal (edges and GWESP statistics). This illustrates the geometrical complexity of the ERGM parameter space with fairly similar ERGM parameters placing weight on vastly different parts of the space of networks. This also demonstrates the difficulty of setting appropriate priors for ERGM parameters. While we could have constrained the space further by choosing a more informative covariance, perhaps based on the single ERGM fits, we decided instead to allow the model fitting algorithm to learn the covariance structure through the data.

To check the prior distribution on the covariance of the window-level parameters, we generated eight sets of networks from a fixed mean $\bar{\mu}^P$ and a covariance generated from the prior. Each set consisted of nine networks, in correspondence with the majority of the observed networks. We compared the summary statistics of these simulated sets of networks with a random sample of eight individuals (Figures 5.5 and 5.6). Even from

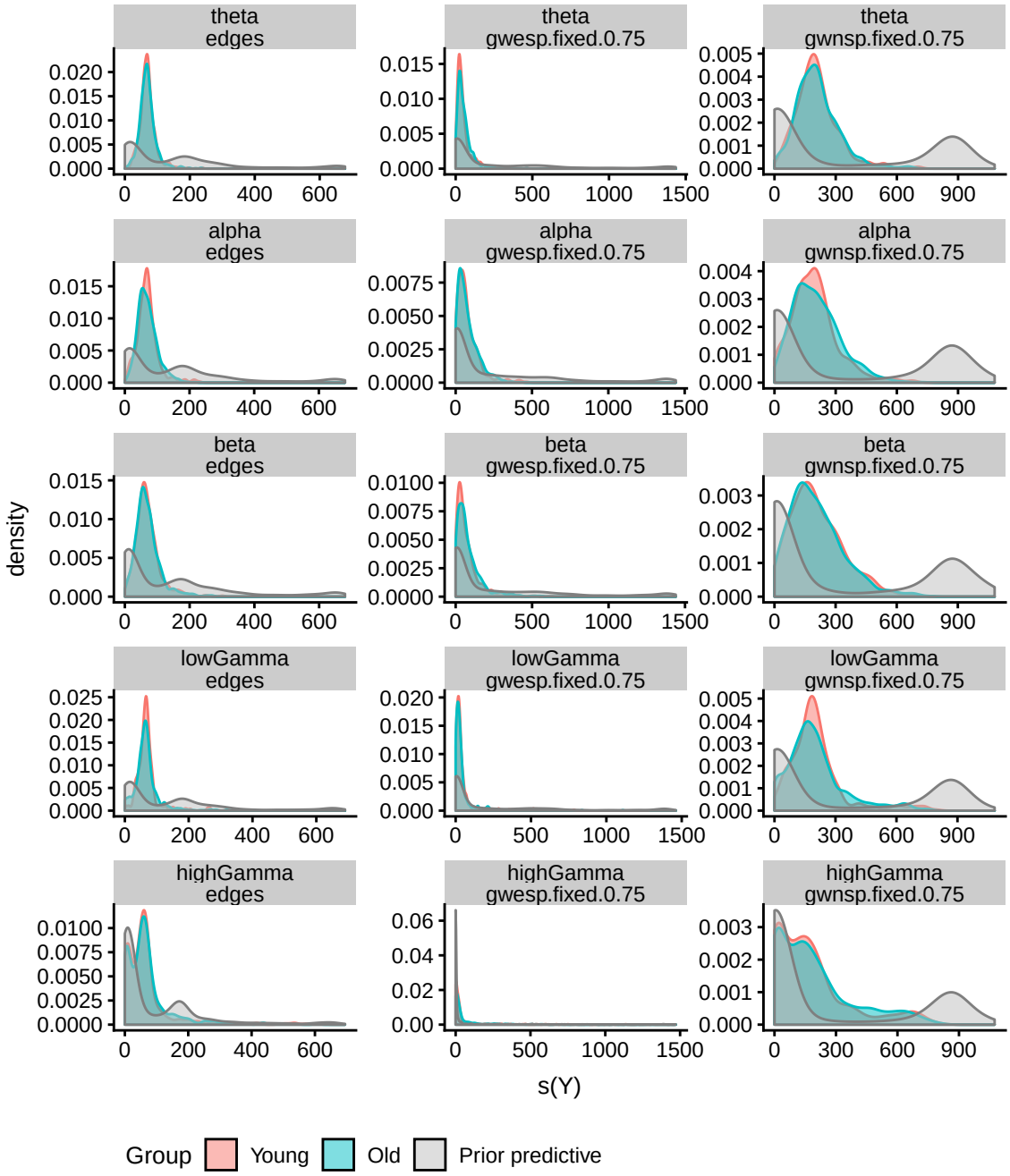


Figure 5.3: Prior predictive check for μ^P - 10% threshold. Here, $S = 2000$ draws from the population-level prior distribution of μ^P were generated. For each draw μ_s^P, we simulated a exponential random graph $\mathbf{y}_s \sim \pi(\cdot|\mu_s^P)$. For each of the frequency bands, the support of the distribution of the resulting summary statistics (grey density plots) covers that of the observed data (red and blue density plots), indicating that the prior is appropriate for this data.

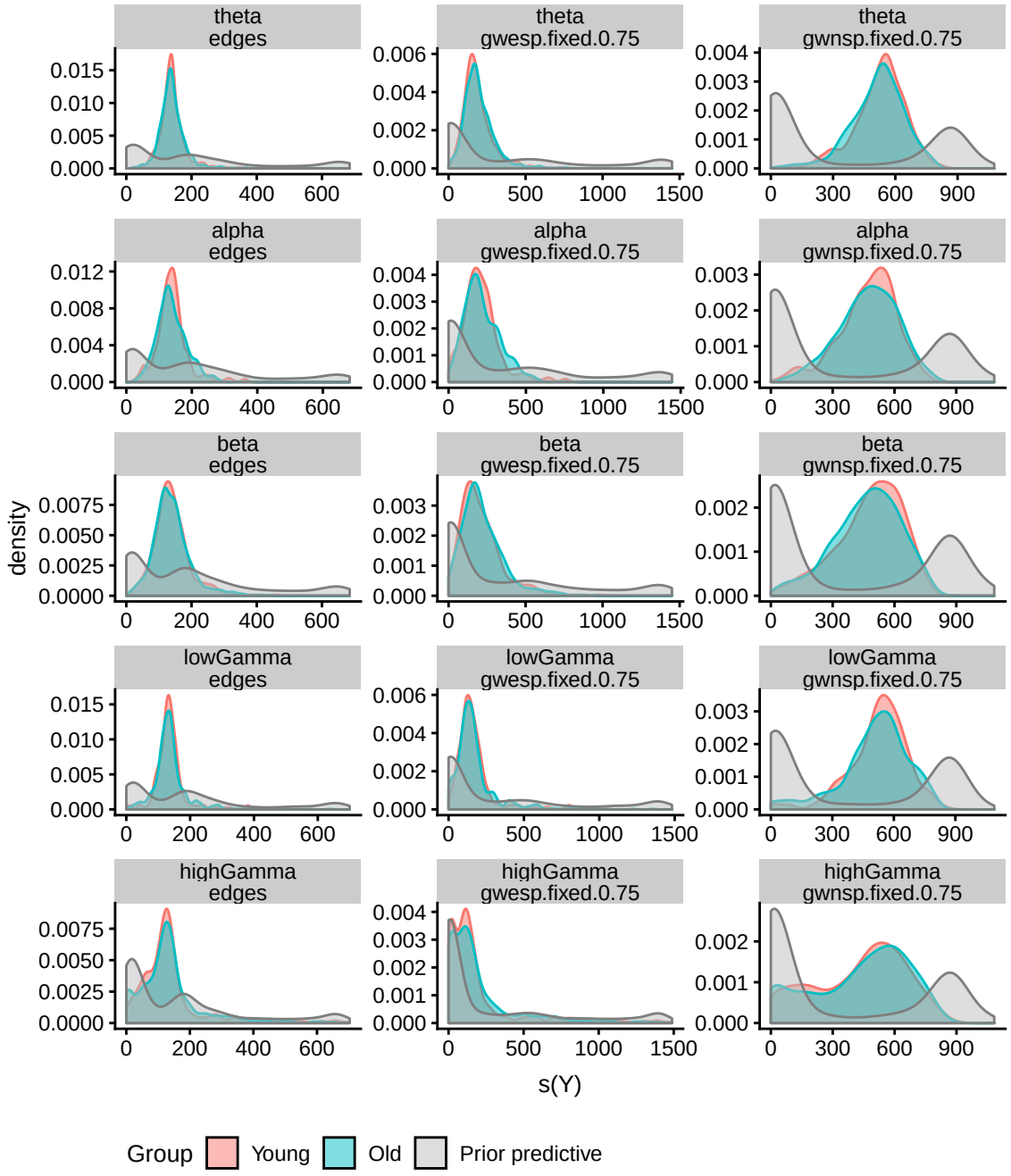


Figure 5.4: Prior predictive check for μ^P - 20% threshold. Here, $S = 2000$ draws from the population-level prior distribution of μ^P were generated. For each draw μ_s^P, we simulated a exponential random graph $\mathbf{y}_s \sim \pi(\cdot|\mu_s^P)$. For each of the frequency bands, the support of the distribution of the resulting summary statistics (grey density plots) covers that of the observed data (red and blue density plots), indicating that the prior is appropriate for this data.

the limited number of examples, we see that the sets of networks simulated from the prior (grey dots) are more diverse than the observed networks (red dots). In particular, the range of the observed sets tends to be small compared to the range of the simulated sets. In a few cases, however, the range of the simulated sets is similar to those observed, indicating that the prior predictive distribution covers the observed data.

Group differences

Although our main focus in this chapter is on assessing differences in the *variability* of functional connectivity, a byproduct of the fitting procedure is posterior samples from the group-level mean parameters $\mu^{(1)}, \mu^{(2)}$, allowing us to compare the overall functional connectivity between groups. Figure 5.7 shows the resulting density plots for the 10% threshold while Figure 5.8 shows the same for the 20% threshold.

Remarkably, the direction of the group differences are largely consistent across the frequency bands and the thresholding value. In each of the ten cases, the posterior samples for the edges parameter tend to be larger for the young group than for the old group. In contrast, the posterior samples for the GWESP parameter tend to be larger for the old group across the runs. The pattern for the GWNSP parameter is less clear. For the 10% threshold, the young group tends to have smaller values except for the β frequency band, where the two groups are largely similar. For the 20% threshold, the GWESP posterior distributions are almost identical across the two groups with the exception of the α band, where the young group tends to have smaller values.

On the other hand, the actual location of each of posterior distributions varies across the frequency bands. For example, the range of the posterior samples of the GWESP parameter across the two groups is about $(0.4, 0.85)$ in the β band and $(0.2, 0.5)$ in the θ band under the 10% threshold. This indicates that, although the direction of the differences between groups are the same, the network structure nevertheless varies across the different frequency bands.

The differences in the determinant of the group-level covariance matrices $\mathbf{\Sigma}^{(1)}, \mathbf{\Sigma}^{(2)}$ are also strikingly consistent across the frequency bands and threshold values (Figure 5.9). For all ten runs except the θ and low γ frequency bands with 20% thresholding, the posterior samples of $|\mathbf{\Sigma}^{(j)}|$ are generally smaller for the old group than for the young group. This indicates that the functional connectivity structure is more variable for individuals in the young group than the old group. Again, the values differ across the frequency bands and threshold values with $|\mathbf{\Sigma}^{(j)}|$ generally increasing with frequency (θ to high γ)

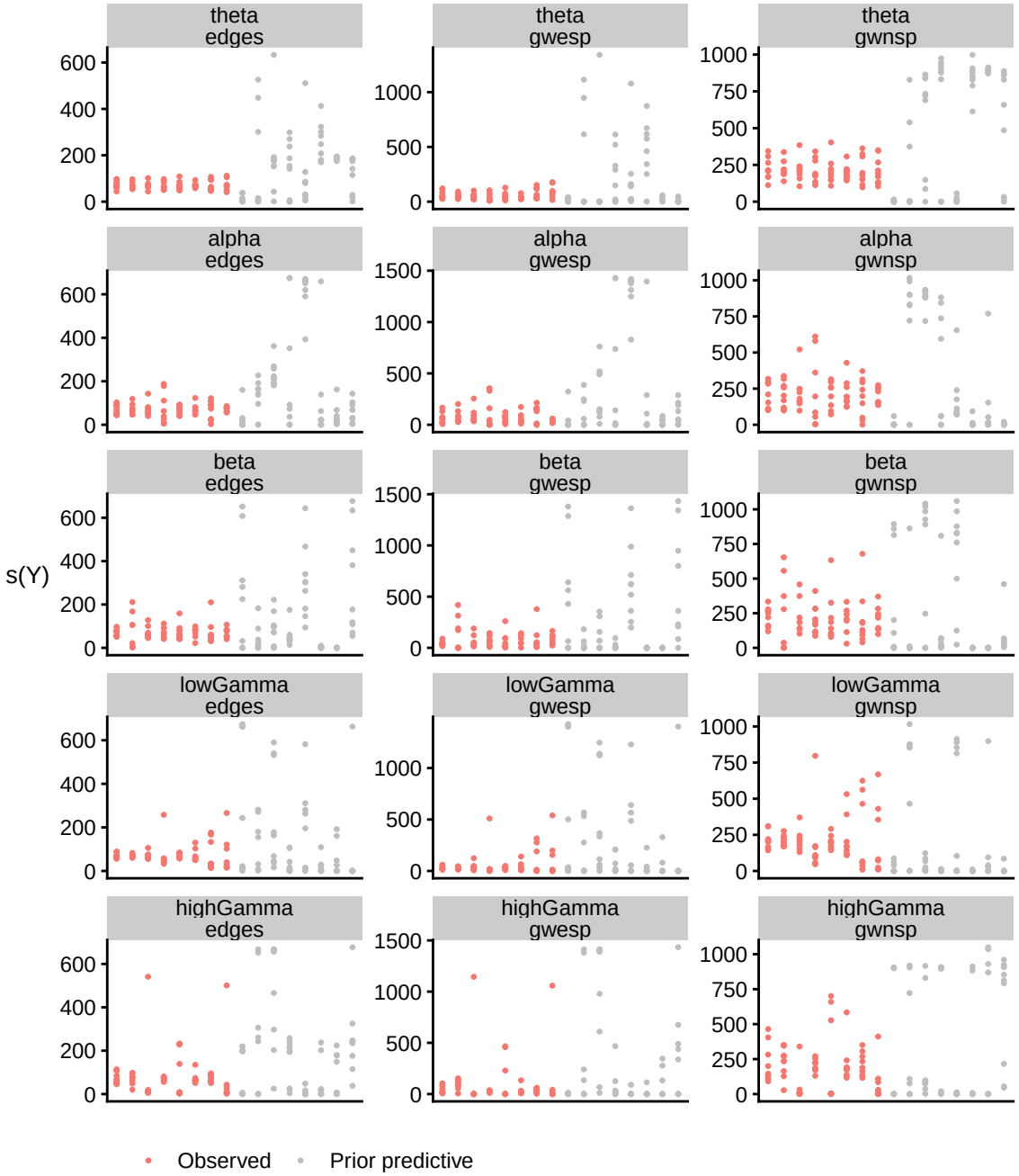


Figure 5.5: Prior predictive check for $\mathbf{\Sigma}_\theta$ - 10% threshold. Here, we generated covariance matrices $\mathbf{\Sigma}_{\theta,s}$ from the prior distribution on $\mathbf{\Sigma}_\theta$ for $s = 1, \ldots, 8$. With each matrix, we then generated a set of $T_s = 9$ ERGM parameters from $\mathcal{N}(\bar{\mu}, \mathbf{\Sigma}_{\theta,s})$ and then simulated a network from each parameter. For each frequency band, this resulted in eight sets of nine networks (grey dots) that we compared with the observed networks of a random subsample of eight individuals (red dots). The range of the simulated sets is much more variable than that of the observed sets, suggesting that the prior predictive distribution covers the observed data.

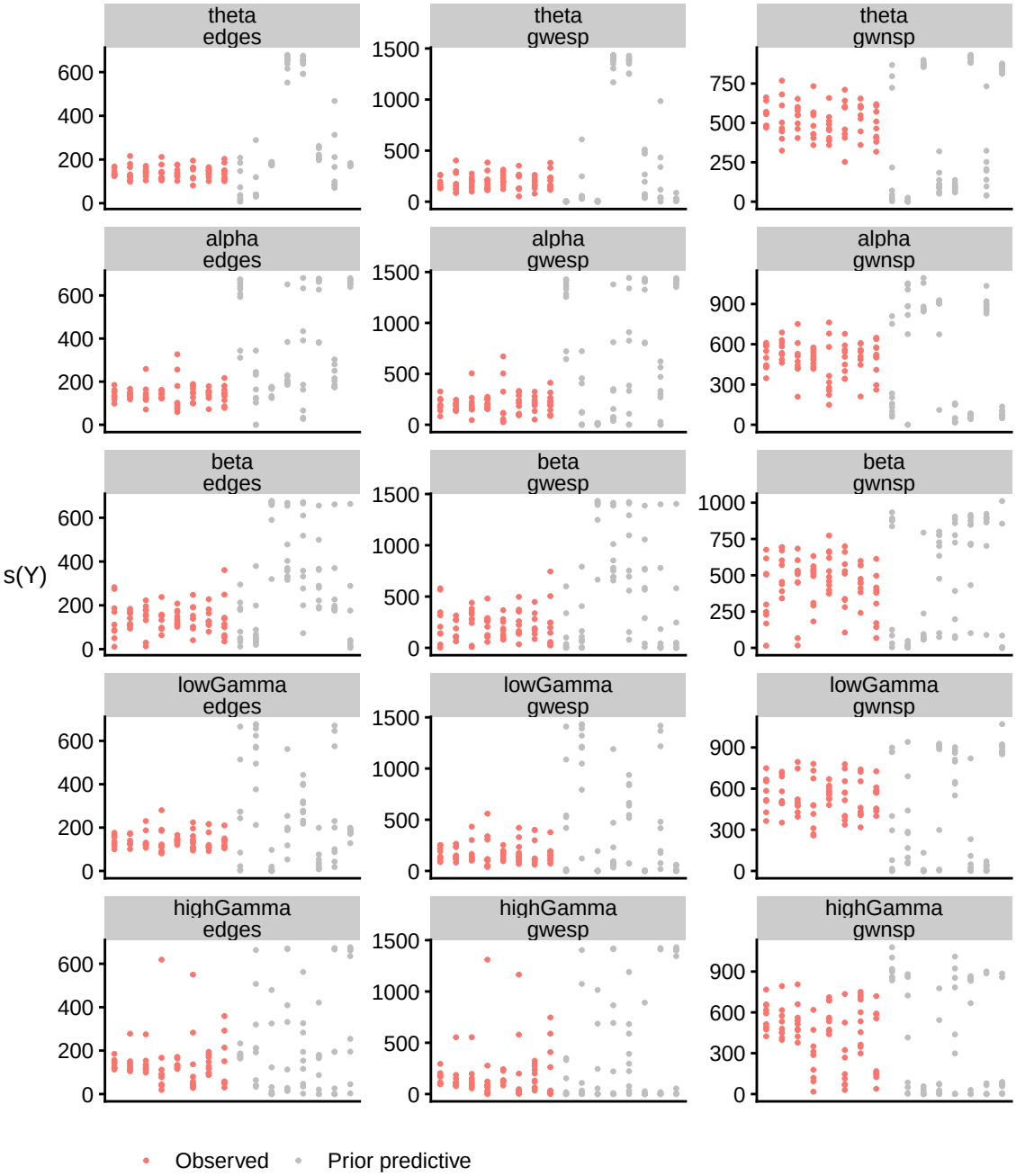


Figure 5.6: Prior predictive check for $\mathbf{\Sigma}_\theta$ - 20% threshold. Here, we generated covariance matrices $\mathbf{\Sigma}_{\theta,s}$ from the prior distribution on $\mathbf{\Sigma}_\theta$ for $s = 1, \ldots, 8$. With each matrix, we then generated a set of $T_s = 9$ ERGM parameters from $\mathcal{N}(\bar{\mu}, \mathbf{\Sigma}_{\theta,s})$ and then simulated a network from each parameter. For each frequency band, this resulted in eight sets of nine networks (grey dots) that we compared with the observed networks of a random subsample of eight individuals (red dots). The range of the simulated sets is much more variable than that of the observed sets, suggesting that the prior predictive distribution covers the observed data.

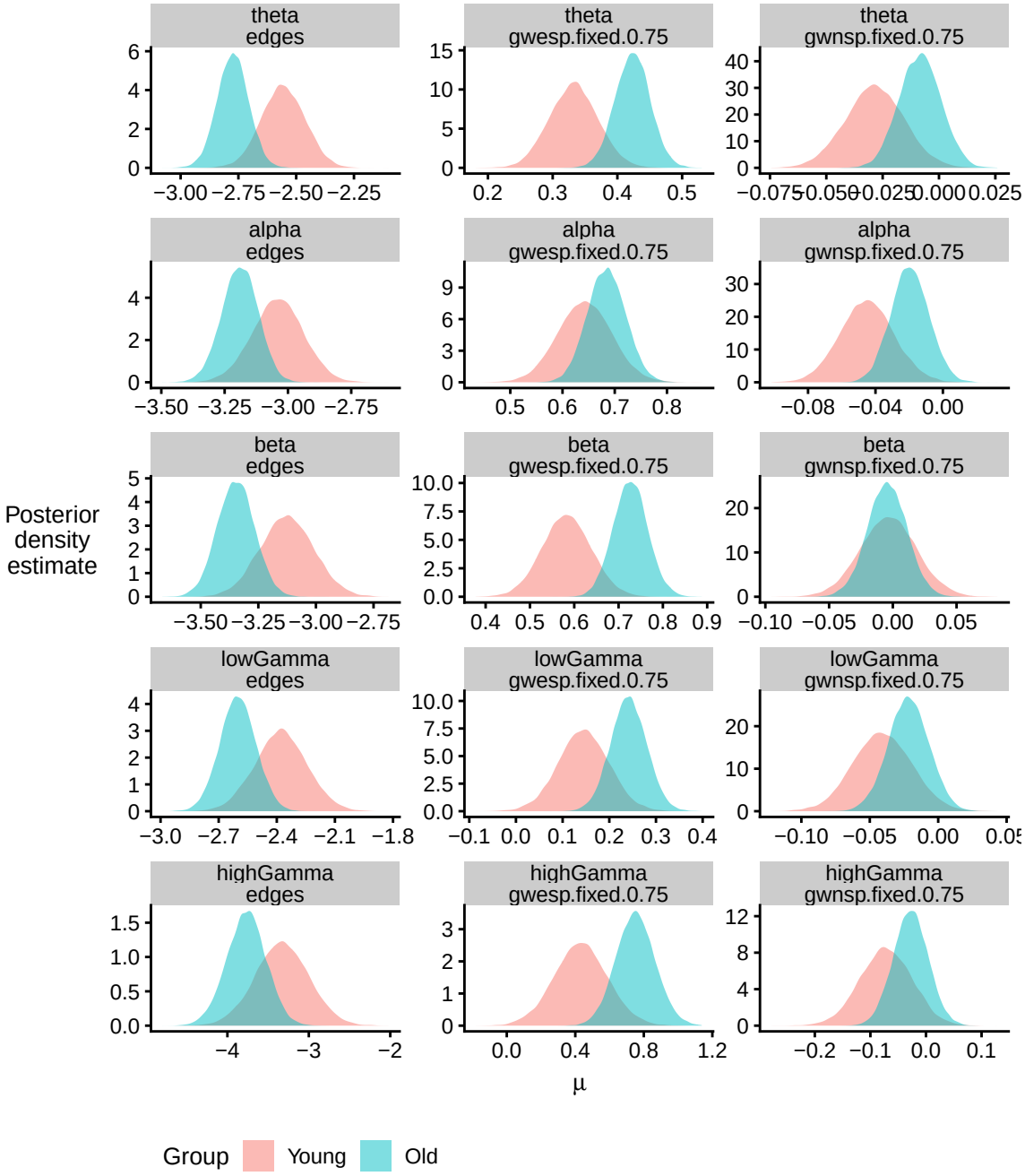


Figure 5.7: Group-level differences in $\mu^{(j)}$ - 10% threshold. The posterior density plots for the group-level means $\mu^{(1)}$ (red plots, young group) and $\mu^{(2)}$ (blue plots, old group) displayed strikingly consistent differences across the frequency bands. For all five frequency bands, the posterior density of the young group for the edges parameter placed more weight on higher values than the young group. This difference was reversed for the GWESP parameter, which tends to be higher for the old group. The differences are not as pronounced for the GWNSP parameter with moderately higher values for the young group across all the frequency bands except β.

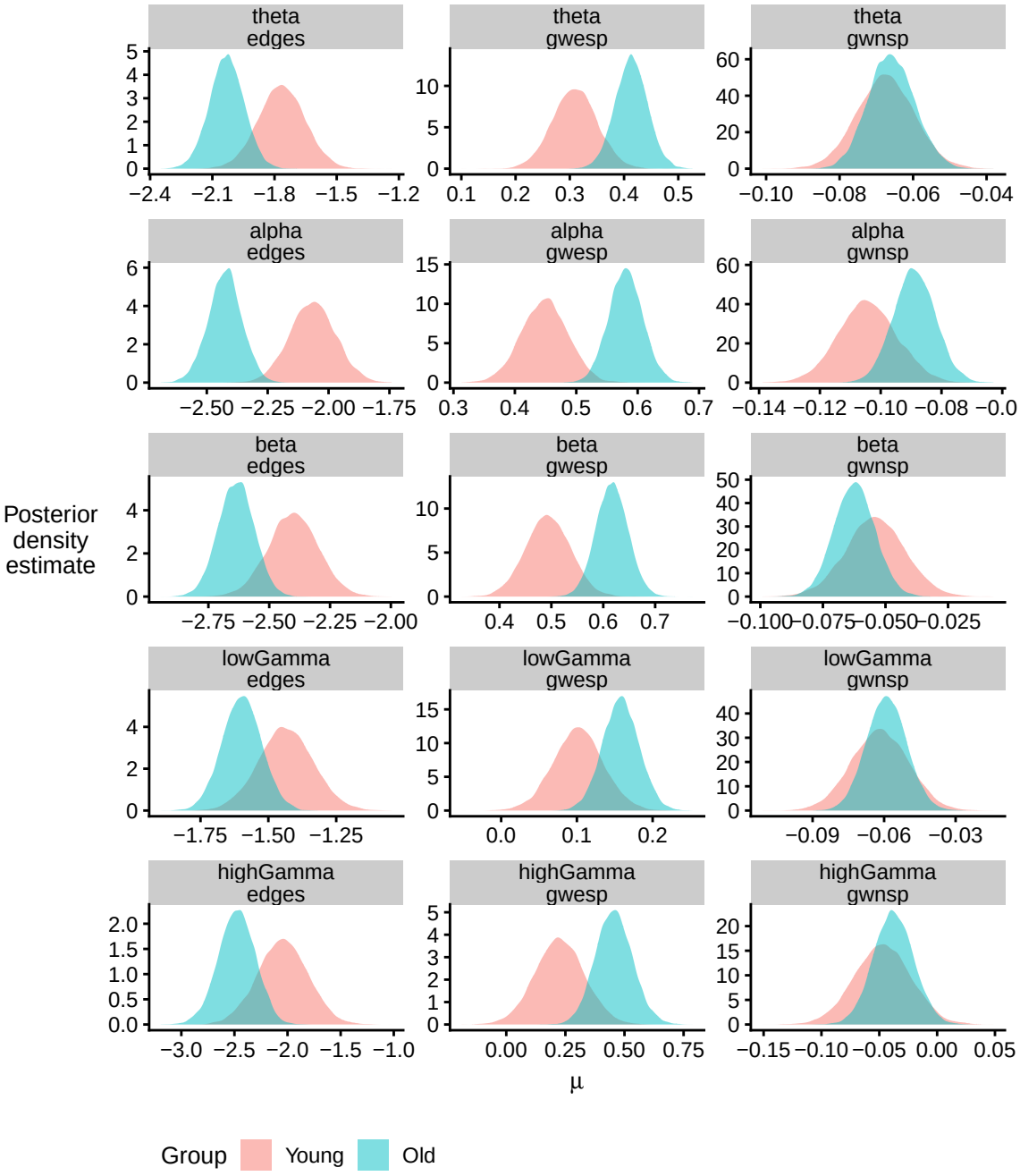


Figure 5.8: Group-level differences in μ - 20% threshold. The posterior density plots for the group-level means $\mu^{(1)}$ (red plots, young group) and $\mu^{(2)}$ (blue plots, old group) displayed strikingly consistent differences across the frequency bands. For all five frequency bands, the posterior density of the young group for the edges parameter placed more weight on higher values than the young group. Except for the α band, there were no clear group differences in the GWNSP parameter.

and smaller for the 20% threshold than the 10% threshold.

Posterior predictive checks

To assess the goodness-of-fit of the model, we performed posterior predictive checks on a random subsample of three individuals per group for each of the runs. Specifically, for each sampled individual i in group j, we selected eight MCMC iterations $k_1, \ldots, k_8 \in \{1, \ldots, K\}$ uniformly at random. Then, for iteration k_s, we drew $T = 9$ window-level parameters $\theta^{(t,i,j)} \sim \mathcal{N}(\theta_{k_s}^{(i,j)}, \mathbf{\Sigma}_{k_s}^{(j)})$, followed by $T = 9$ exponential random graphs from $\mathbf{y}^{(t,i,j)'} \sim \pi(\cdot|\theta^{(t,i,j)})$. We then compared the summary statistics of these simulated sets of networks with the observed data for each sample individual (Figures 5.10 to 5.14).

For each of the runs, the range of the summary statistics of the networks simulated from the posterior predictive distribution was generally closer to the observed data than the range of those simulated from the prior predictive distribution (see Figures 5.5 and 5.6). For example, the range of edges was typically around (0,600) for sets of networks simulated from the prior predictive distribution but only at most (0, 400) for sets simulated from the posterior predictive distribution in the θ, α and β frequency bands.

For the θ, α and β frequency bands, the range of the simulated network summary statistics appeared to be the same or greater than the spread of the observed networks for each individual, suggesting that the posterior distributions of the group-level covariance matrices $\mathbf{\Sigma}^{(1)}, \mathbf{\Sigma}^{(2)}$ overestimated the variance between windows for each individual. On the other hand, for both the γ bands, the spread tended to be the same if not smaller for the simulated sets of networks. This was usually due to a single network that has much larger values of summary statistics than the remaining networks for that individual - see, for example, individual 2 in Figure 5.13a.

5.5 Discussion

In this chapter, we have presented how the multilevel framework introduced in Chapters 2 and 3 can be extended to the case of multiple networks per individual. This is highly relevant in the analysis of dynamic functional connectivity, which can be characterised as a sequence of networks across the course of a brain scan. The extension amounts to adding another level of hierarchy to the framework: *window* - individual - group - population.

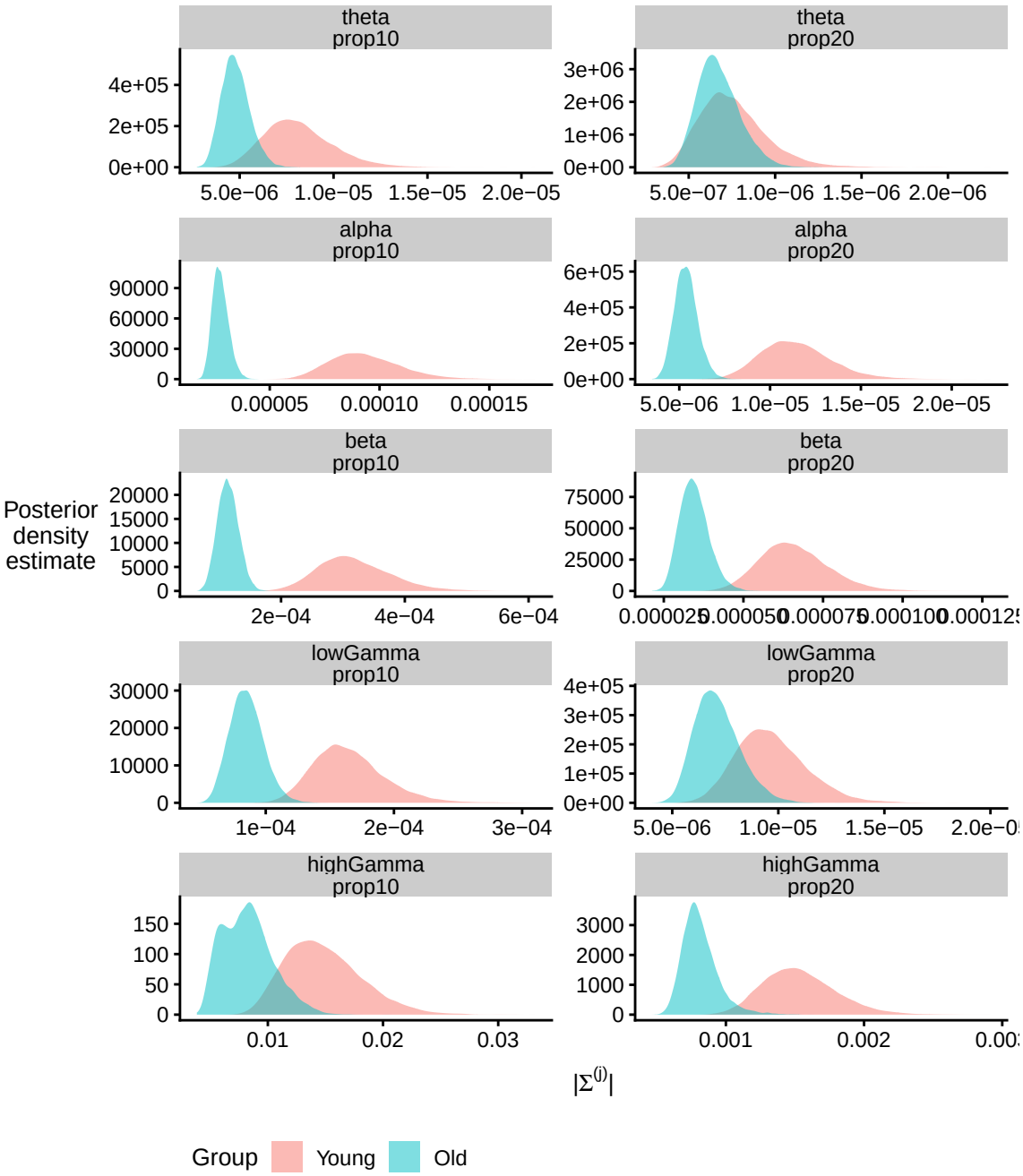
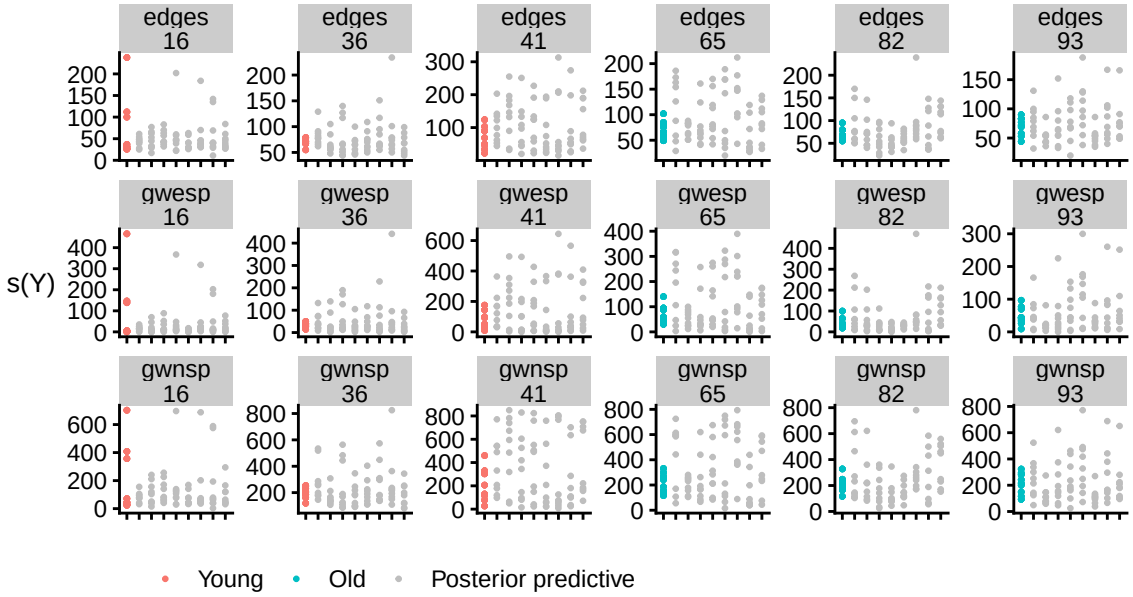


Figure 5.9: Group-level differences in $|\mathbf{\Sigma}^{(j)}|$. The posterior density plots for $|\mathbf{\Sigma}^{(1)}|$ (red plots, young group) and $|\mathbf{\Sigma}^{(2)}|$ (blue plots, old group) again show consistent differences across the five frequency bands and the two threshold values. With the exception of the θ band with the 20% threshold (top right plot), the posterior samples of the covariance matrix for the young group are generally larger than those of the old group, indicating a more variable connectivity structure.

(a) 10% threshold



(b) 20% threshold

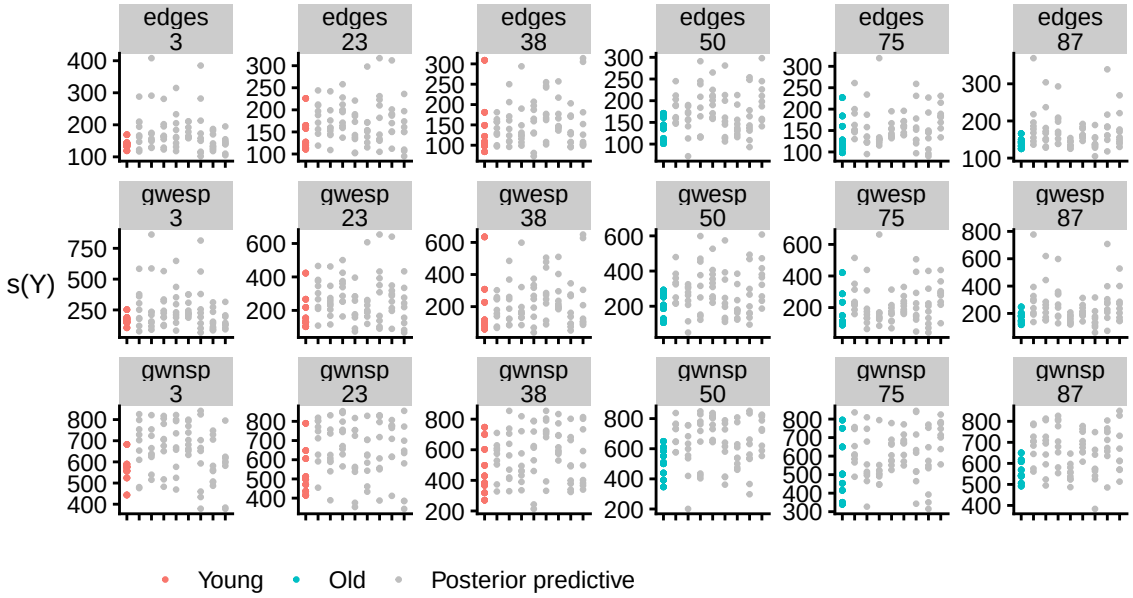
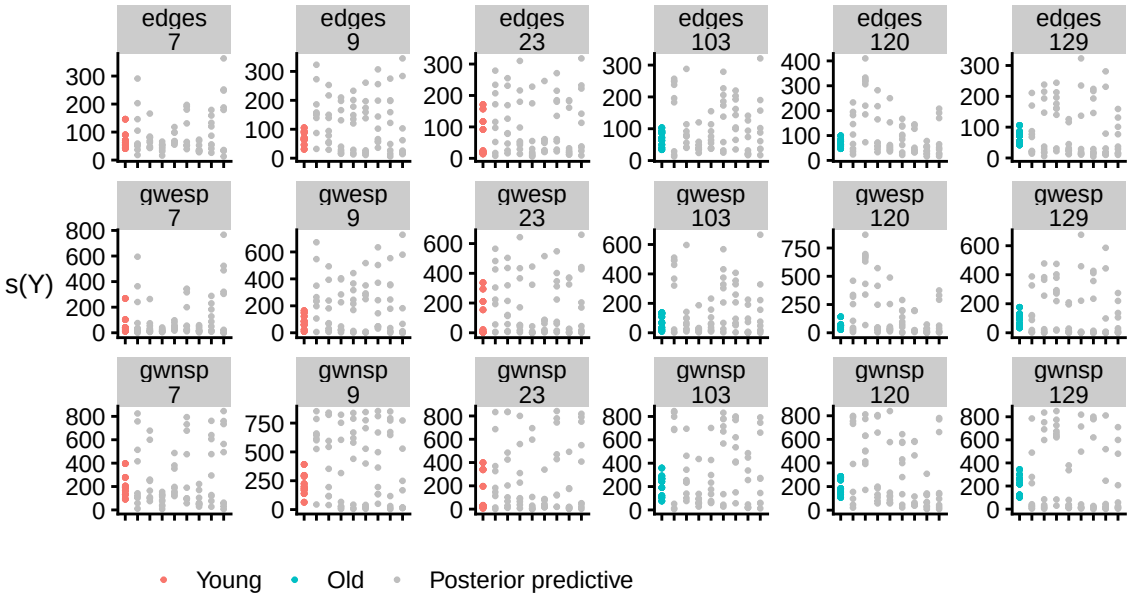


Figure 5.10: Posterior predictive check for the θ frequency band. The numbers below the summary statistics refer to the different sampled individuals. For each sampled individual in both thresholds, there was at least two sets of networks simulated from the posterior predictive distribution (grey dots) that had broadly similar summary statistics to those observed (coloured dots), indicating a reasonable fit to the data. While some of the simulated sets had a much larger range than the observed data, this was nevertheless smaller than the range of the sets simulated from the prior predictive distribution (Figure 5.5, top row).

(a) 10% threshold



(b) 20% threshold

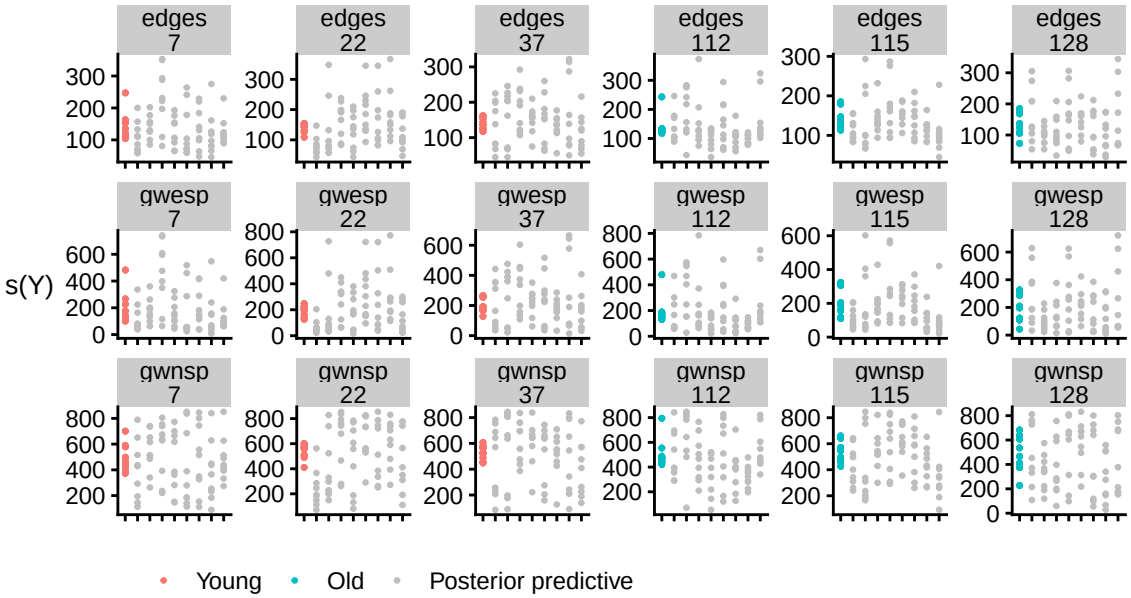
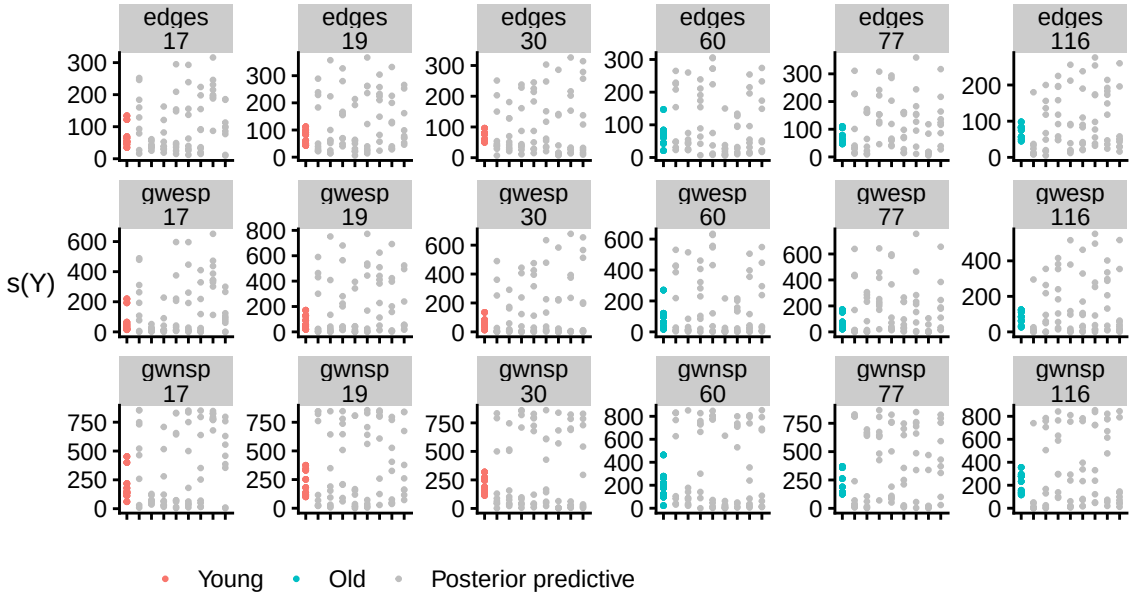


Figure 5.11: Posterior predictive check for the α frequency band. The numbers below the summary statistics refer to the different sampled individuals. With the exception of individual 9 for the 10% threshold and individual 22 for the 20% threshold, the posterior predictive distribution indicated a reasonable fit with at two or three of the simulated sets (grey dots) matching the observed data (coloured dots) closely.

(a) 10% threshold



(b) 20% threshold

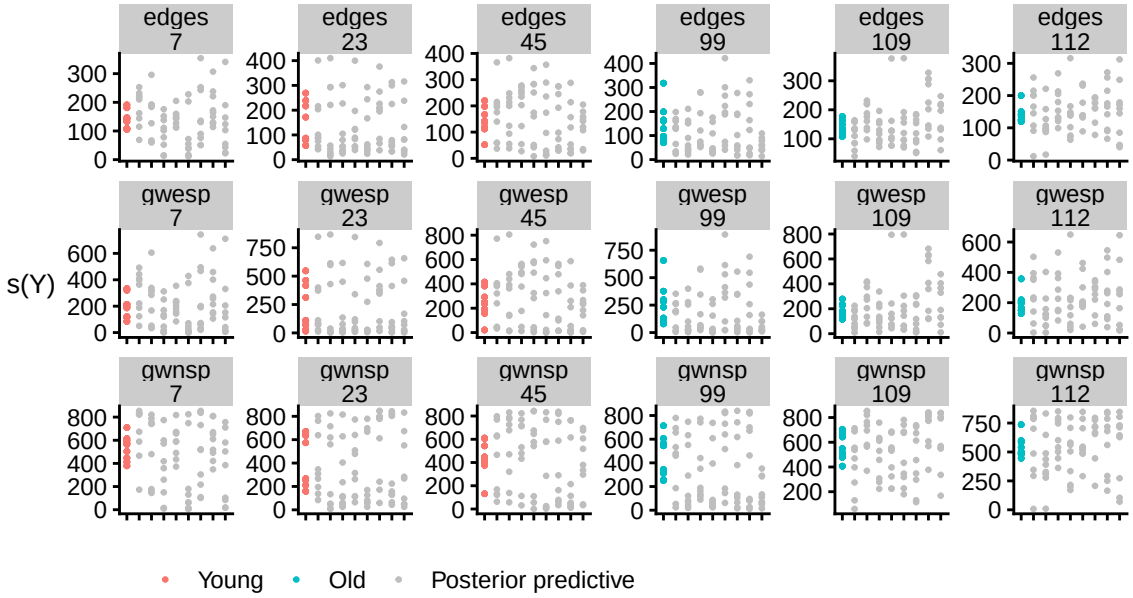
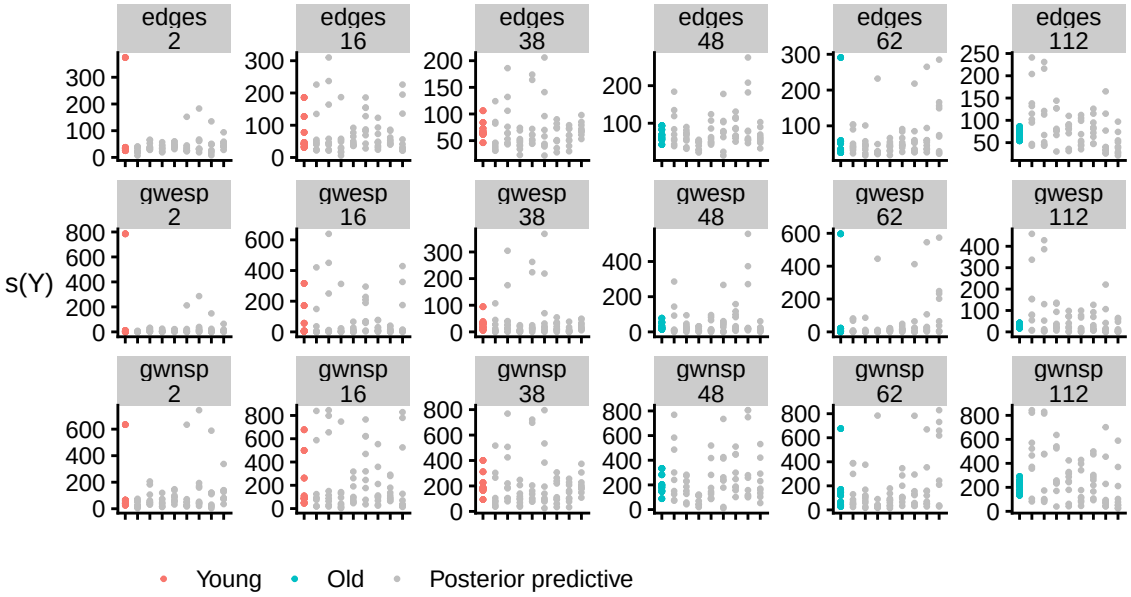


Figure 5.12: Posterior predictive check for the β frequency band. The numbers below the summary statistics refer to the different sampled individuals. a) For the 10% threshold, the sets of networks simulated from the posterior predictive distribution (grey dots) generally had a larger range than those of the observed sets of networks for both the young individuals (red dots) and the old individuals (blue dots), indicating that the posterior overestimated the variability of networks across windows. b) On the other hand, the model was a better fit for sets of networks constructed with the 20% threshold with the simulated sets displaying a more similar range to the observed sets.

(a) 10% threshold



(b) 20% threshold

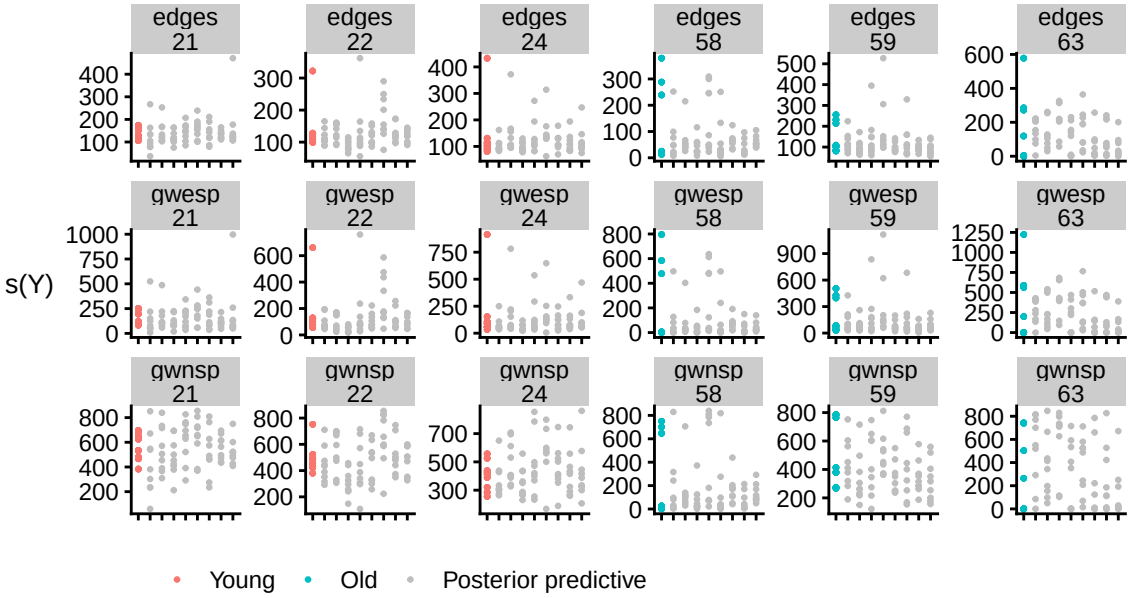
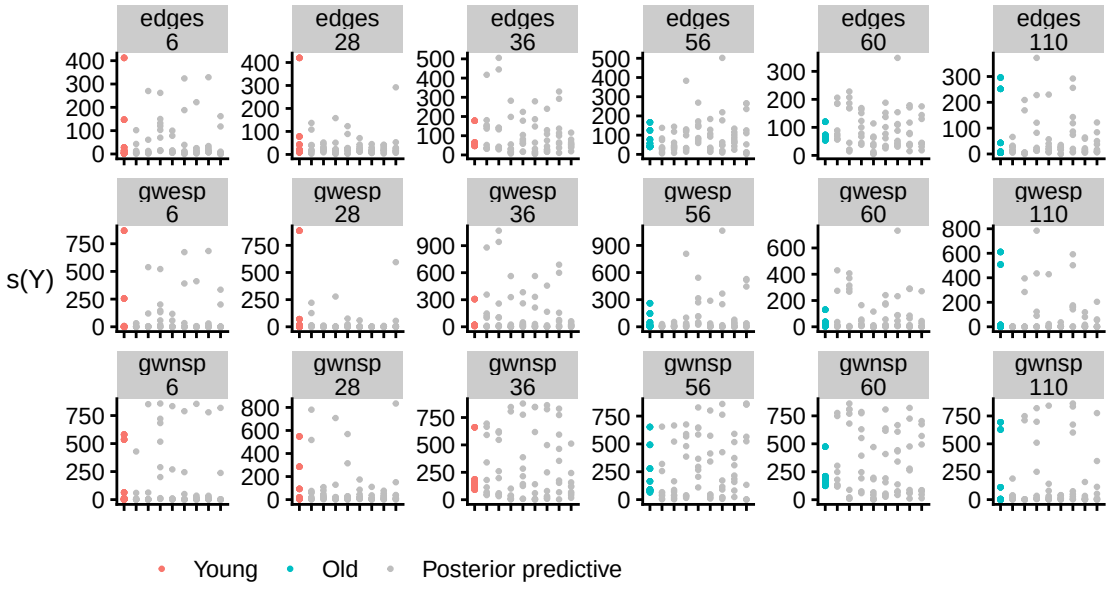


Figure 5.13: Posterior predictive check for the low γ frequency band. The numbers below the summary statistics refer to the different sampled individuals. With the exception of individual 2 with the 10% threshold (top left) and individual 63 with the 20% threshold (bottom right), the summary statistics of sets of networks simulated from the posterior predictive distribution (grey dots) were broadly comparable with those of the observed sets (coloured dots). For the two aforementioned individuals, both had a single network with a large value for edges and GWESP, that was not captured in any of the corresponding simulated sets of networks.

(a) 10% threshold



(b) 20% threshold

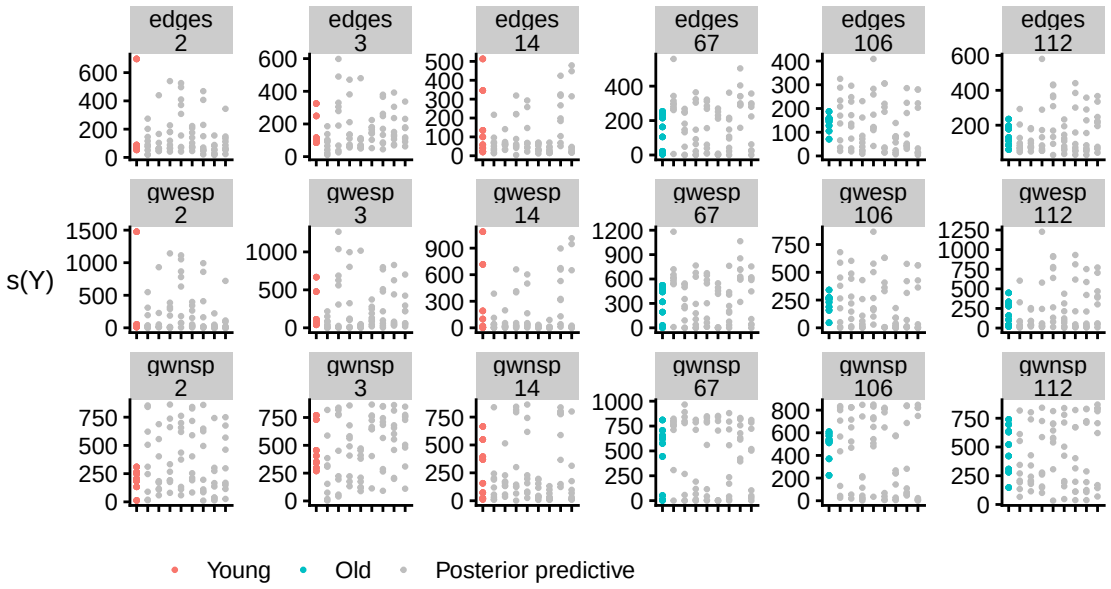


Figure 5.14: Posterior predictive check for the high γ frequency band. The numbers below the summary statistics refer to the different sampled individuals. With the exception of individual 6 with the 10% threshold (top left) and individual 2 with the 20% threshold (bottom left), the summary statistics of sets of networks simulated from the posterior predictive distribution (grey dots) were broadly comparable with those of the observed sets (coloured dots). Under the 10% threshold, individual 6 had a single network with a large value for edges and GWESP, that was not captured in any of the corresponding simulated sets of networks. For individual 2 under the 20% threshold, the range of the GWESP statistic for each of the simulated sets of networks was much larger than that of the observed data.

As with the case of static functional connectivity, the framework allows the pooling of information across individuals within the same group and across groups. By setting up the model appropriately and analysing the resulting posterior samples produced by the MCMC algorithm, we can determine whether there are differences in the both the overall connectivity structure and the variability in connectivity structure between a group of young individuals and a group of old individuals.

Differences in connectivity structure are consistent across frequency bands

Although the parameter estimates of both the group-level means and the determinant of the group-level covariance matrices varied across the two thresholding values and the five frequency bands, the direction of the differences between the young group and the old group remained strikingly consistent. For the group-level means, the GWESP parameter (corresponding broadly to local efficiency) was generally higher in the old group than the young group. This was offset by lower estimates of the edges parameter, while differences in the GWNSP were not as pronounced.

The determinant of the group-level covariance matrix, a measure of variability in connectivity structure, was generally larger for the young group than for the old group. This is in line with previous studies that have found variation in functional connectivity to decrease with age [28, 165]. Admittedly, the covariance matrix determinant is a rather broad measure of variability and more work is needed to understand precisely which aspects of the network structure are more variable across groups.

Comparison to fMRI results

In Chapter 3, we performed a static connectivity analysis using fMRI data to compare the functional connectivity structure between the young group and the old group. The dynamic connectivity analysis performed here provided an alternative way to characterise the group-level static functional connectivity structure via the group-level mean parameters. While the network construction methods in the fMRI analyses (absolute and proportional thresholding) differed to that used in this Chapter (individual-based thresholding), we saw some similarities across the imaging modalities. In particular, the group differences under the absolute 20% threshold for the fMRI networks appear to be similar to those under the 20% threshold for the MEG networks. These results should be treated with caution, however, given the poor fit of this model for the fMRI data, and further work is needed to understand which network features persist across modalities.

Goodness-of-fit

The assessment of model adequacy is challenging due to 1) the hierarchical structure of the model and 2) the difficulties in comparing populations of networks. As we were predominantly interested in detecting differences in the variability of functional connectivity structure, we performed posterior predictive checks based on the covariance matrices governing the spread of the window-level parameters. For each of the thresholds and frequency bands, we compared the summary statistics of the networks of a random subsample of the individuals in the two groups with sets of networks simulated from the posterior predictive distribution. A small number of the observed sets of networks did not appear to be captured well by the posterior predictive distribution, indicating a poor fit.

This may have been due to the simplifying assumption of a common group-level covariance matrix governing the spread of window-level parameters. Although a more complex model with individual-level covariance matrices may have been more realistic, it is likely that estimates of these would have been unreliable given the small number of window-level parameters ($T \approx 9$) for each individual. One possible modification to increase the reliability of such individual-level covariance matrices would be to use smaller windows, resulting in more networks per individual, though this would come at an increased computational cost. Another possibility would be to add a layer of hierarchy, allowing information to be pooled across individuals at the level of the covariance matrices.

5.5.1 Limitations

Lack of temporal dependence

One important assumption of the model is that, for each individual, the window-level parameters are drawn independently from the respective individual-level multivariate normal distribution, thus omitting any potential for temporal dependence from one window to the next. This is unlikely to be a realistic assumption, given that functional connectivity has been shown to vary with arousal levels, which themselves vary during the course of a scan [45]. In the presence of overlapping windows, which would allow assessing variations in functional connectivity at a finer temporal resolution, this assumption would unquestionably be violated given that consecutive networks are estimated from (partially) the same data.

Computational cost

While the static functional connectivity application consisted of one network per individual, the number of networks considered here increased by a factor of about ten, corresponding to the number of windows in the sliding-window analysis. The total number of networks fit (for each frequency band and both thresholding values) was 1224 so each MCMC iteration required 2448 ERGM simulations. This was by far the most computationally intensive part of the fitting procedure. Accordingly, the time taken to fit the algorithm was substantial: approximately six days for the longest run.

Given the high temporal resolution of MEG, it is generally possible to estimate dynamic functional connectivity using shorter, possibly overlapping, windows. For example, in a task-based study, [105] used windows of width $w = 6s$ and steps of $\tau = 0.5s$, resulting in hundreds of networks per individual. The current computational cost of the fitting procedure would be a clear limiting factor of the framework when applied to data of this scale. As discussed in Section 2.3, however, this could be mitigated by running the ERGM simulations in parallel given access to a sufficient number of computing cores.

5.5.2 Future work

The key, and limiting, assumption of the model presented in this chapter is the lack of temporal dependence from one time window's network to the next. There are at least two avenues towards relaxing this assumption that warrant further investigation. Firstly, a time-varying coefficient exponential random graph model [88] has been developed to characterise the smooth evolution of a single temporal network's topology over time. This could be extended for a population of temporal networks, though some care would be needed to appropriately specify the multilevel structure across networks.

An alternative approach would be to incorporate state-based methods such as the hidden Markov model. These have previously been used with alternative characterisations of functional connectivity such as autoregressive processes [158] and stochastic blockmodels [119] to describe variations in connectivity structure. In the context of our framework, one could associate each window with a hidden state, and model changes in connectivity via hidden state transitions. This additional complexity, however, would make inference even more challenging and would likely require a novel approach to model fitting.

Chapter 6

Conclusion

The main contribution of this thesis is to introduce a novel framework for modelling populations of networks. This framework, based on exponential random graphs, uses the machinery of Bayesian hierarchical modelling to pool information across networks. We have applied this method to brain functional connectivity networks constructed from resting-state fMRI data from the Cam-CAN project, a study on healthy ageing, to investigate age-related changes. We then shifted our focus to dynamic functional connectivity data, performing a simulation study to understand how dFC can be affected by various sources of heterogeneity between individuals. Finally, we extended our framework to handle dynamic functional connectivity and applied this extension to resting-state MEG data. We now discuss some of the challenges facing our framework as well as some avenues for future work.

6.1 Challenges

While exponential random graph models are a flexible class of statistical distributions for network data, they carry with them a number of issues that need to be addressed before our framework can be applied more successfully. Note that these challenges also face ERGMs in the (standard) case of single networks, and are magnified in the presence of multiple networks.

Selection of summary statistics

A key element of the model is the choice of network summary statistics. We used the summary statistics suggested in Simpson et al. [132], namely: number of edges, geometrically-weighted edgewise shared partners statistic, and geometrically-weighted non-edgewise shared partners statistic. Although the fitting algorithm appeared to converge, the subsequent posterior predictive checks indicated that the model fit was poor. This suggests that the choice of these summary statistics was not appropriate for our data.

The choice of summary statistics is a challenging task. Most existing statistics were developed in the context of social networks, and it is likely that novel metrics are needed to better characterise brain networks. While we applied the same summary statistics for both fMRI and MEG as well as each thresholding procedure, it may be that different sets of summary statistics are appropriate in each of the different cases. Further work is needed to develop new summary statistics and to investigate which combinations of summary statistics can parsimoniously describe brain connectivity structure.

Methods to select summary statistics for ERGMs of single networks typically rely on some notion of goodness-of-fit. The most common approach is to use a graphical goodness-of-fit test, comparing the observed network to networks simulated according to the fitted ERGM on a selection of graph metrics deemed to be important [76]. There are some drawbacks to this approach. The choice of which network metrics to compare on is somewhat subjective: a model may appear adequate under one set of metrics but not under another. Although it may be obvious when a model is not adequate for the data, it may not be clear which summary statistics should be added or removed in order to improve the fit. Further, for populations of networks, there is an added layer of difficulty in visually comparing the observed data to data generated from the fitted model.

Quantitative approaches to model selection also exist. For example, the spectral goodness-of-fit measure [130] uses properties of the graph Laplacian to quantify the difference between an observed network and networks simulated from the fitted ERGM. Bayesian model selection for ERGMs has also been developed, using a reversible-jump MCMC procedure adapted to handle the intractable ERGM likelihood [23]. This approach, however, is highly computationally intensive and may not be practical when extended to deal with multiple networks simultaneously.

Interpretation of parameters

The parameter values of an ERGM quantify the relative influence of the respective summary statistic on the overall network structure. Specifically, each individual component can be interpreted as the difference in conditional log-odds per unit change in the corresponding summary statistic between two networks that differ by one edge. *All else being equal*, a larger value of a given component places more weight on networks with a larger value of the corresponding summary statistic.

In practice, however, summary statistics tend to be correlated. For example, consider an ERGM with two summary statistics: number of edges and number of triangles. A network with a larger number of triangles is likely to have a larger number of edges. When comparing one network with the same number of edges but more triangles than a second network, the ERGM fitted to the first network will have a larger value for the triangles parameter. Importantly, this will be accompanied by a lower value for the edges parameter, even though the number of edges are the same for the two networks. While it may be straightforward to interpret models with only two summary statistics, this reasoning becomes more difficult when faced with three or more summary statistics that may likely have non-linear relationships with each other.

Given these difficulties, we advocate mapping the fitted ERGM parameter values back to the scale of the summary statistics or, more generally, other network metrics. The posterior predictive checks we performed indicate one possible approach: simulate a large number of populations of networks from the fitted model parameters and compare, for example, the population median of the summary statistics. For a more complete assessment, one should also compare other properties of the distribution of the summary statistics such as the extrema and the range.

Computational efficiency

The intractable normalising constant in the ERGM likelihood poses the main barrier to inference in our proposed framework. To circumvent this, we developed a novel MCMC procedure, combining the exchange algorithm [102], adapted for Bayesian ERGMs [24], with Gibbs sampling. The main computational bottleneck lies in simulating the exponential random graphs for the exchange step at each MCMC iteration. While the computational cost of our fitting procedure should increase roughly linearly in the number of networks, Krivitsky & Handcock [86] used an empirical approach to show that the

cost may grow on the order of $p(|V| + |E|)\log(|E|)$ where p is the number of summary statistics, $|V|$ is the number of nodes, and $|E|$ is the number of edges. This restricts the size of the networks that our framework can handle in practical time to about 100 nodes. As a result, the performance of our fitting procedure will benefit greatly from any improvements in the computational efficiency of exponential random graph simulation.

6.2 Future work

Our framework can be applied to any population of networks. Although we focused on (dynamic) functional connectivity, the framework could be used, for example, to characterise anatomical connectivity for a group of individuals. The flexibility of both exponential random graph models and Bayesian hierarchical modelling also offer some promising opportunities to develop our framework in a number of directions.

Multimodal data

Recently, there has been increasing interest in combining distinct neuroimaging modalities (such as fMRI and MEG) to provide new insights into brain function [151]. For connectivity analyses, our framework provides a natural way of combining information from different imaging technologies. Networks constructed from different modalities can be modelled simultaneously as exponential random graphs, with the relationships between networks specified via the multilevel framework.

Alternatively, one could incorporate information regarding structural connectivity via the specification of the prior. The anatomical structure of the brain is believed to play a central role in supporting the brain's functional connectivity structure [154]. By first fitting our framework to structural data, one could construct an informative prior from the resulting information to constrain the space of functional networks to those supported by the brain's anatomy.

ERGM extensions

There have been a number of recent extensions to the standard exponential random graph model, which could be applied within our multilevel framework. For example, to avoid the issue of using an arbitrary thresholding procedure, one could use weighted exponential random graph models [84, 38], extending the binary ERGM to weighted networks.

Various temporal ERGMs have been proposed that could explicitly model the temporal properties of functional connectivity [85, 88]. The hierarchical ERGM (not to be confused with our multilevel framework) allow for local dependence in networks, as opposed to the global dependence induced by conventional ERGMs, and thus can describe more complex network structures and may be more amenable to statistical inference [124, 125]. While each of these extensions bring with them added complexity that needs to be accounted for in the fitting algorithm, they offer a wide range of tools that could be used to model populations of diverse types of networks.

Bibliography

[1] Chapter 3 - connectivity matrices and brain graphs. In Alex Fornito, Andrew Zalesky, and Edward T. Bullmore, editors, *Fundamentals of Brain Network Analysis*, pages 89 – 113. Academic Press, San Diego, 2016.

[2] Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and Ed Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006.

[3] Howard J. Aizenstein, Kristi A. Clark, Meryl A. Butters, Jennifer Cochran, V. Andrew Stenger, Carolyn C. Meltzer, Charles F. Reynolds, and Cameron S. Carter. The bold hemodynamic response in healthy aging. *Journal of Cognitive Neuroscience*, 16(5):786–793, Jun 2004.

[4] Andrew L. Alexander, Jee Eun Lee, Mariana Lazar, and Aaron S. Field. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4(3):316–329, Jul 2007.

[5] Aaron Alexander-Bloch, Nitin Gogtay, David Meunier, Rasmus Birn, Liv Clasen, Francois Lalonde, Rhoshel Lenroot, Jay Giedd, and Edward Bullmore. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in Systems Neuroscience*, 4:147, 2010.

[6] Thomas W Allan, Susan T Francis, Cesar Caballero-Gaudes, Peter G Morris, Elizabeth B Liddle, Peter F Liddle, Matthew J Brookes, and Penny A Gowland. Functional connectivity in mri is driven by spontaneous bold events. *PloS one*, 10(4):e0124577, 2015.

[7] Elena A. Allen, Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 2012.

[8] Micah Allen, Davide Poggiali, Kirstie Whitaker, Tom R Marshall, and Rogier Kievit. Raincloud plots: a multi-platform tool for robust data visualization. *PeerJ Preprints*, 6:e27137v1, 2018.

[9] P. Alquier, N. Friel, R. Everitt, and A. Boland. Noisy monte carlo: convergence of markov chains with approximate transition kernels. *Statistics and Computing*, 26(1):29–47, Jan 2016.

[10] Mohammad R Arbabshirani, Eduardo Castro, and Vince D Calhoun. Accurate classification of schizophrenia patients based on novel resting-state fmri features. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6691–6694. IEEE, 2014.

[11] Mohammad R Arbabshirani, Eswar Damaraju, Ronald Phlypo, Sergey Plis, Elena Allen, Sai Ma, Daniel Mathalon, Adrian Preda, Jatin G Vaidya, Tülay Adali, et al. Impact of autocorrelation on functional connectivity. *Neuroimage*, 102:294–308, 2014.

[12] F Gregory Ashby. *Statistical analysis of fMRI data*. MIT press, 2011.

[13] Sylvain Baillet. Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, 20(3):327–339, March 2017.

[14] Danielle S. Bassett, Mason A. Porter, Nicholas F. Wymbs, Scott T. Grafton, Jean M. Carlson, and Peter J. Mucha. Robust detection of dynamic community structure in networks. *Chaos*, 23(1):013142, Mar 2013. 022301CHA[PII].

[15] Danielle S. Bassett, Nicholas F. Wymbs, Mason A. Porter, Peter J. Mucha, Jean M. Carlson, and Scott T. Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641–7646, 2011.

[16] Danielle S Bassett, Nicholas F Wymbs, M Puck Rombach, Mason A Porter, Peter J Mucha, and Scott T Grafton. Task-based core-periphery organization of human brain dynamics. *PLoS Comput Biol*, 9(9):e1003171, 2013.

[17] André M. Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in Systems Neuroscience*, 9:175, 2016.

[18] Shankar Bhamidi, Guy Bresler, and Allan Sly. Mixing time of exponential random graphs. *Ann. Appl. Probab.*, 21(6):2146–2170, 12 2011.

[19] Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995.

[20] Steven L. Bressler and Vinod Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6):277 – 290, 2010.

[21] Matthew J. Brookes, Joanne R. Hale, Johanna M. Zumer, Claire M. Stevenson, Susan T. Francis, Gareth R. Barnes, Julia P. Owen, Peter G. Morris, and Srikantan S. Nagarajan. Measuring functional connectivity using meg: Methodology and comparison with fcmri. *NeuroImage*, 56(3):1082 – 1104, 2011.

[22] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198, Mar 2009.

[23] A. Caimo and N. Friel. Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11 – 24, 2013.

[24] Alberto Caimo and Nial Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41 – 55, 2011.

[25] Alberto Caimo and Nial Friel. Bergm: Bayesian exponential random graphs in r. *Journal of Statistical Software, Articles*, 61(2):1–25, 2014.

[26] Catie Chang and Gary H. Glover. Time–frequency dynamics of resting-state brain connectivity measured with fmri. *NeuroImage*, 50(1):81 – 98, 2010.

[27] Ming-Hui Chen. Importance-weighted marginal bayesian posterior density estimation. *Journal of the American Statistical Association*, 89(427):818–824, 1994.

[28] Yuanyuan Chen, Weiwei Wang, Xin Zhao, Miao Sha, Ya'nan Liu, Xiong Zhang, Jianguo Ma, Hongyan Ni, and Dong Ming. Age-related decline in the variation of dynamic functional connectivity: A resting state analysis. *Frontiers in Aging Neuroscience*, 9:203, 2017.

[29] Siddhartha Chib. Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321, 1995.

[30] G.L. Colclough, M.J. Brookes, S.M. Smith, and M.W. Woolrich. A symmetric multivariate leakage correction for meg connectomes. *NeuroImage*, 117:439 – 448, 2015.

[31] G.L. Colclough, M.W. Woolrich, P.K. Tewarie, M.J. Brookes, A.J. Quinn, and S.M. Smith. How reliable are meg resting-state connectivity metrics? *NeuroImage*, 138:284 – 293, 2016.

[32] Nicolas A. Crossley, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T. Fox, Philip McGuire, and Edward T. Bullmore. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395, 2014.

[33] Rhodri Cusack, Alejandro Vicente-Grabovetsky, Daniel J. Mitchell, Conor J. Wild, Tibor Auer, Annika C. Linke, and Jonathan E. Peelle. Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using matlab and xml. *Frontiers in Neuroinformatics*, 8:90, 2015.

[34] Mandeep S. Dagli, John E. Ingeholm, and James V. Haxby. Localization of cardiac-induced signal change in fmri. *NeuroImage*, 9(4):407 – 415, 1999.

[35] Madelaine Daianu, Neda Jahanshad, Talia M. Nir, Arthur W. Toga, Clifford R. Jack, Michael W. Weiner, and Paul M. Thompson, for the Alzheimer's Disease Neuroimaging Initiative. Breakdown of brain connectivity between normal aging and alzheimer's disease: A structural k-core network analysis. *Brain Connectivity*, 3(4):407–422, 2013. PMID: 23701292.

[36] John Dell'Italia, Micah A. Johnson, Paul M. Vespa, and Martin M. Monti. Network analysis in disorders of consciousness: Four problems and one proposed solution (exponential random graph models). *Frontiers in Neurology*, 9:439, 2018.

[37] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31(3):968 – 980, 2006.

[38] Bruce A Desmarais and Skyler J Cranmer. Statistical inference for valued-edge networks: The generalized exponential random graph model. *PloS one*, 7(1):e30136, 2012.

[39] Mark D'Esposito, Eric Zarahn, Geoffrey K. Aguirre, and Bart Rypma. The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. *NeuroImage*, 10(1):6 – 14, 1999.

[40] Eugene P. Duff, Tamar Makin, Michiel Cottaar, Stephen M. Smith, and Mark W. Woolrich. Disambiguating brain functional connectivity. *NeuroImage*, 173:540 – 550, 2018.

[41] Daniele Durante, David B Dunson, et al. Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis*, 2016.

[42] Amanda Elton and Wei Gao. Task-related modulation of functional connectivity variability and its behavioral correlations. *Human brain mapping*, 36(8):3260–3272, 2015.

[43] Richard G. Everitt. Bayesian parameter estimation for latent markov random fields and social networks. *Journal of Computational and Graphical Statistics*, 21(4):940–960, 2012.

[44] Fabrizio De Vico Fallani, Vito Latora, and Mario Chavez. A Topological Criterion for Filtering Information in Complex Brain Networks. *PLOS Computational Biology*, 13(1):e1005305, January 2017.

[45] Jin Fan, Pengfei Xu, Nicholas T. Van Dam, Tehila Eilam-Stock, Xiaosi Gu, Yue-jia Luo, and Patrick R. Hof. Spontaneous brain activity relates to autonomic arousal. *Journal of Neuroscience*, 32(33):11176–11186, 2012.

[46] Karl J. Friston. Functional and Effective Connectivity: A Review. *Brain Connectivity*, 1(1):13–36, January 2011.

[47] K.J. Friston, J.T. Ashburner, S.J. Kiebel, W.D. Penny, and T.E. Nichols. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press. Elsevier Science, 2007.

[48] K.J. Friston, O. Josephs, E. Zarahn, A.P. Holmes, S. Rouquette, and J.-B. Poline. To smooth or not to smooth?: Bias and efficiency in fmri time-series analysis. *NeuroImage*, 12(2):196 – 208, 2000.

[49] Joaquín M. Fuster. The Module: Crisis of a Paradigm. *Neuron*, 26(1):51 – 53, 2000.

[50] Joaquín M. Fuster. The cognit: A network model of cortical representation. *International Journal of Psychophysiology*, 60(2):125 – 132, 2006.

[51] Linda Geerligs, Cam-CAN, and Richard N. Henson. Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. *NeuroImage*, 135:16 – 31, 2016.

[52] Linda Geerligs, Natasha M. Maurits, Remco J. Renken, and Monicque M. Lorist. Reduced specificity of functional connectivity in the aging brain during task performance. *Human Brain Mapping*, 35(1):319–330, 2014.

[53] Linda Geerligs, Mikail Rubinov, Cam-CAN, and Richard N. Henson. State and trait components of functional connectivity: Individual differences vary with mental state. *The Journal of Neuroscience*, 35(41):13949–13961, 2015.

[54] Linda Geerligs, Kamen A. Tsvetanov, Cam-CAN, and Richard N. Henson. Challenges in measuring individual differences in functional connectivity using fmri: The case of healthy aging. *Human Brain Mapping*, 38(8):4125–4156, 2017.

[55] Alan E. Gelfand, Sujit K. Sahu, and Bradley P. Carlin. Efficient parametrisations for normal linear mixed models. *Biometrika*, 82(3):479–488, 1995.

[56] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2013.

[57] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.

[58] Zoubin Ghahramani, Matthew J Beal, et al. Variational inference for bayesian mixtures of factor analysers. In *NIPS*, volume 12, pages 449–455, 1999.

[59] W. R. Gilks, G. O. Roberts, and E. I. George. Adaptive direction sampling. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1):179–189, 1994.

[60] Cedric E. Ginestet, Arnaud P. Fournel, and Andrew Simmons. Statistical network analysis for functional mri: summary networks and group comparisons. *Frontiers in Computational Neuroscience*, 8:51, 2014.

[61] Cedric E. Ginestet, Jun Li, Prakash Balachandran, Steven Rosenberg, and Eric D. Kolaczyk. Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.*, 11(2):725–750, 06 2017.

[62] Cedric E. Ginestet, Thomas E. Nichols, Ed T. Bullmore, and Andrew Simmons. Brain network analysis: Separating cost from topology using cost-integration. *PLOS ONE*, 6(7):1–17, 07 2011.

[63] Gary H. Glover. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, 22(2):133 – 139, 2011. Functional Imaging.

[64] Joseph V. Hajnal, Ralph Myers, Angela Oatridge, Jane E. Schwieso, Ian R. Young, and Graeme M. Bydder. Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magnetic Resonance in Medicine*, 31(3):283–291, 1994.

[65] Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.*, 65:413–497, Apr 1993.

[66] Alexander Hammers, Richard Allom, Matthias J. Koepp, Samantha L. Free, Ralph Myers, Louis Lemieux, Tejal N. Mitchell, David J. Brooks, and John S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, 19(4):224–247, 2003.

[67] Mark S Handcock. Assessing degeneracy in statistical models of social networks. Technical report, Citeseer, 2003.

[68] Daniel A. Handwerker, Vinai Roopchansingh, Javier Gonzalez-Castillo, and Peter A. Bandettini. Periodic changes in fmri connectivity. *NeuroImage*, 63(3):1712 – 1719, 2012.

[69] Faith M Hanlon, Nicholas A Shaff, Andrew B Dodd, Josef M Ling, Juan R Bustillo, Christopher C Abbott, Shannon F Stromberg, Swala Abrams, Denise S Lin, and Andrew R Mayer. Hemodynamic response function abnormalities in schizophrenia during a multisensory detection task. *Human brain mapping*, 37(2):745–755, 2016.

[70] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[71] Joerg F Hipp, David J Hawellek, Maurizio Corbetta, Markus Siegel, and Andreas K Engel. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature neuroscience*, 15(6):884–890, 2012.

[72] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97 – 125, 2012. Temporal Networks.

[73] Scott A. Huettel, Jeffrey D. Singerman, and Gregory McCarthy. The effects of aging upon the hemodynamic response measured by functional {MRI}. *NeuroImage*, 13(1):161 – 175, 2001.

[74] David Hunter, Mark Handcock, Carter Butts, Steven Goodreau, and Martina Morris. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software, Articles*, 24(3):1–29, 2008.

[75] David R. Hunter. Curved exponential family models for social networks. *Social Networks*, 29(2):216 – 230, 2007. Special Section: Advances in Exponential Random Graph (p*) Models.

[76] David R Hunter, Steven M Goodreau, and Mark S Handcock. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258, 2008.

[77] R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H. Duyn, Gary H. Glover, Javier Gonzalez-Castillo, Daniel A. Handwerker, Shella Keilholz, Vesa Kiviniemi, David A. Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, and Catie Chang. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, 80:360 – 378, 2013. Mapping the Connectome.

[78] David T Jones, Prashanthi Vemuri, Matthew C Murphy, Jeffrey L Gunter, Matthew L Senjem, Mary M Machulda, Scott A Przybelski, Brian E Gregg, Kejal Kantarci, David S Knopman, et al. Non-stationarity in the "resting brain's" modular architecture. *PloS one*, 7(6):e39731, 2012.

[79] Fikret Işik Karahanoğlu and Dimitri Van De Ville. Transient brain activity disentangles fmri resting-state dynamics in terms of spatially and temporally overlapping networks. *Nature communications*, 6, 2015.

[80] Vesa Kiviniemi, Tapani Vire, Jukka Remes, Ahmed Abou Elseoud, Tuomo Starck, Osmo Tervonen, and Juha Nikkinen. A sliding time-window ica reveals spatial variability of the default mode network in time. *Brain Connectivity*, 1(4):339–347, Sep 2011.

[81] Eric D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer Series in Statistics. Springer-Verlag, New York, 2009.

[82] JH Koskinen. Bayesian analysis of exponential random graphs-estimation of parameters and model selection. Technical report, Stockholm University, 2004.

[83] Johan H. Koskinen, Garry L. Robins, Peng Wang, and Philippa E. Pattison. Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks*, 35(4):514–527, October 2013.

[84] Pavel N. Krivitsky. Exponential-family random graph models for valued networks. *Electron. J. Statist.*, 6:1100–1128, 2012.

[85] Pavel N. Krivitsky and Mark S. Handcock. A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46, 2014.

[86] Pavel N. Krivitsky and Mark S. Handcock. Supplementary material: A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46, 2014.

[87] Pavel N. Krivitsky, Mark S. Handcock, and Martina Morris. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, 8(4):319 – 339, 2011.

[88] Jihui Lee, Gen Li, and James D Wilson. Varying-coefficient models for dynamic networks. *arXiv preprint arXiv:1702.03632*, 2017.

[89] B.C.L. Lehmann, S.R. White, R.N. Henson, Cam-CAN, and L. Geerligs. Assessing dynamic functional connectivity in heterogeneous samples. *NeuroImage*, 157:635 – 647, 2017.

[90] Nora Leonardi and Dimitri Van De Ville. On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage*, 104:430 – 436, 2015.

[91] Kaiming Li, Lei Guo, Jingxin Nie, Gang Li, and Tianming Liu. Review of methods for functional brain connectivity detection using fmri. *Computerized Medical Imaging and Graphics*, 33(2):131–139, 2009.

[92] Jun S. Liu and Ying Nian Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274, 1999.

[93] Xiao Liu and Jeff H Duyn. Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, 110(11):4392–4397, 2013.

[94] Fernando Lopes da Silva. EEG and MEG: Relevance to Neuroscience. *Neuron*, 80(5):1112–1128, December 2013.

[95] Torben E. Lund. fcmri—mapping functional connectivity or correlating cardiac-induced noise? *Magnetic Resonance in Medicine*, 46(3):628–628, 2001.

[96] Clarisse I. Mark, Erin L. Mazerolle, and J. Jean Chen. Metabolic and vascular origins of the bold effect: Implications for imaging pathology and resting-state brain function. *Journal of Magnetic Resonance Imaging*, 42(2):231–246, 2015.

[97] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[98] David Meunier, Sophie Achard, Alexa Morcom, and Ed Bullmore. Age-related changes in modular organization of human brain functional networks. *NeuroImage*, 44(3):715 – 723, 2009.

[99] Robyn L. Miller, Maziar Yaesoubi, Jessica A. Turner, Daniel Mathalon, Adrian Preda, Godfrey Pearlson, Tulay Adali, and Vince D. Calhoun. Higher dimensional meta-state analysis reveals reduced resting fmri connectivity dynamism in schizophrenia patients. *PLOS ONE*, 11(3):1–24, 03 2016.

[100] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

[101] Sophia Mueller, Danhong Wang, Michael D. Fox, B.T. Thomas Yeo, Jorge Sepulcre, Mert R. Sabuncu, Rebecca Shafee, Jie Lu, and Hesheng Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586 – 595, 2013.

[102] Iain Murray, Zoubin Ghahramani, and David J. C. MacKay. MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pages 359–366. AUAI Press, 2006.

[103] Mark Newman. *Networks*. Oxford University Press, Oxford, New York, new edition, second edition edition, July 2018.

[104] Catalina Obando and Fabrizio De Vico Fallani. A statistical model for brain networks inferred from large-scale electrophysiological signals. *Journal of The Royal Society Interface*, 14(128):20160940, March 2017.

[105] George C. O'Neill, Prejaas K. Tewarie, Giles L. Colclough, Lauren E. Gascoyne, Benjamin A.E. Hunt, Peter G. Morris, Mark W. Woolrich, and Matthew J. Brookes. Measurement of dynamic task related functional networks using {MEG}. *NeuroImage*, 146:667 – 678, 2017.

[106] Omiros Papaspiliopoulos, Gareth O. Roberts, and Martin Sköld. Non-centered parameterisations for hierarchical models and data augmentation. In JM Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, and M West, editors, *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, volume 307. Oxford University Press, USA, 2003.

[107] Omiros Papaspiliopoulos, Gareth O. Roberts, and Martin Sköld. A general framework for the parametrization of hierarchical models. *Statist. Sci.*, 22(1):59–73, 02 2007.

[108] Ameera X. Patel, Prantik Kundu, Mikail Rubinov, P. Simon Jones, Petra E. Vértes, Karen D. Ersche, John Suckling, and Edward T. Bullmore. A wavelet method for modeling and despiking motion artifacts from resting-state fmri time series. *NeuroImage*, 95:287 – 304, 2014.

[109] Tiago P. Peixoto. Reconstructing networks with unknown and heterogeneous errors. *Phys. Rev. X*, 8:041011, Oct 2018.

[110] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006.

[111] Russell A. Poldrack. Region of interest analysis for fmri. *Social Cognitive and Affective Neuroscience*, 2(1):67–70, 2007.

[112] Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59(3):2142–2154, 2012.

[113] Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, 160:41 – 54, 2017. Functional Architecture of the Brain.

[114] D. Rangaprakash, Guo-Rong Wu, Daniele Marinazzo, Xiaoping Hu, and Gopikrishna Deshpande. Hemodynamic response function (hrf) variability confounds resting-state fmri functional connectivity. *Magnetic Resonance in Medicine*, 80(4):1697–1713, 2018.

[115] Alessandro Rinaldo, Stephen E. Fienberg, and Yi Zhou. On the geometry of discrete exponential families with application to exponential random graph models. *Electron. J. Statist.*, 3:446–484, 2009.

[116] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statist. Sci.*, 16(4):351–367, 11 2001.

[117] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173 – 191, 2007. Special Section: Advances in Exponential Random Graph (p*) Models.

[118] Garry Robins, Tom Snijders, Peng Wang, Mark Handcock, and Philippa Pattison. Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29(2):192 – 215, 2007. Special Section: Advances in Exponential Random Graph (p*) Models.

[119] Lucy F. Robinson, Lauren Y. Atlas, and Tor D. Wager. Dynamic functional connectivity using state-based dynamic community structure: Method and application to opioid analgesia. *NeuroImage*, 108:274 – 291, 2015.

[120] Baxter P. Rogers, Victoria L. Morgan, Allen T. Newton, and John C. Gore. Assessing functional connectivity in the human brain by fmri. *Magnetic Resonance Imaging*, 25(10):1347 – 1357, 2007.

[121] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059 – 1069, 2010. Computational Models of the Brain.

[122] Ünal Sakoğlu, Godfrey D Pearlson, Kent A Kiehl, Y Michelle Wang, Andrew M Michael, and Vince D Calhoun. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(5-6):351–366, 2010.

[123] Daniel L. Schacter, Randy L. Buckner, Wilma Koutstaal, Anders M. Dale, and Bruce R. Rosen. Late onset of anterior prefrontal activity during true and false recognition: An event-related fmri study. *NeuroImage*, 6(4):259 – 269, 1997.

[124] Michael Schweinberger and Mark S. Handcock. Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):647–676, 2015.

[125] Michael Schweinberger and Pamela Luna. hergm: Hierarchical exponential-family random graph models. *Journal of Statistical Software, Articles*, 85(1):1–39, 2018.

[126] James G Scott, Ryan C Kelly, Matthew A Smith, Pengcheng Zhou, and Robert E Kass. False discovery rate regression: an application to neural synchrony detection in primary visual cortex. *Journal of the American Statistical Association*, 110(510):459–471, 2015.

[127] Meredith A. Shafto, Lorraine K. Tyler, Marie Dixon, Jason R. Taylor, James B. Rowe, Rhodri Cusack, Andrew J. Calder, William D. Marslen-Wilson, John Duncan, Tim Dalgleish, Richard N. Henson, Carol Brayne, and Fiona E. Matthews. The cambridge centre for ageing and neuroscience (cam-can) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, 14(1):1–25, 2014.

[128] Sadia Shakil, Chin-Hui Lee, and Shella Dawn Keilholz. Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states. *NeuroImage*, 133:111 – 128, 2016.

[129] James M. Shine, Oluwasanmi Koyejo, Peter T. Bell, Krzysztof J. Gorgolewski, Moran Gilat, and Russell A. Poldrack. Estimation of dynamic functional connectivity using multiplication of temporal derivatives. *NeuroImage*, 122:399 – 407, 2015.

[130] Jesse Shore and Benjamin Lubin. Spectral goodness of fit for network models. *Social Networks*, 43:16 – 27, 2015.

[131] Sean Simpson, Robert Lyday, Satoru Hayasaka, Anthony Marsh, and Paul Laurienti. A permutation testing framework to compare groups of brain networks. *Frontiers in Computational Neuroscience*, 7:171, 2013.

[132] Sean L Simpson, Satoru Hayasaka, and Paul J Laurienti. Exponential random graph modeling for complex brain networks. *PloS one*, 6(5):e20039, 2011.

[133] Sean L. Simpson, Malaak N. Moussa, and Paul J. Laurienti. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *NeuroImage*, 60(2):1117 – 1126, 2012.

[134] Michel R.T. Sinke, Rick M. Dijkhuizen, Alberto Caimo, Cornelis J. Stam, and Willem M. Otte. Bayesian exponential random graph modeling of whole-brain structural networks across lifespan. *NeuroImage*, 135(Supplement C):79 – 91, 2016.

[135] Ann E. Sizemore and Danielle S. Bassett. Dynamic graph metrics: Tutorial, toolbox, and tale. *NeuroImage*, 180:417 – 427, 2018. Brain Connectivity Dynamics.

[136] Anne M. Smith, Bobbi K. Lewis, Urs E. Ruttimann, Frank Q. Ye, Teresa M. Sinnwell, Yihong Yang, Jeff H. Duyn, and Joseph A. Frank. Investigation of low frequency drift in fmri signal. *NeuroImage*, 9(5):526 – 533, 1999.

[137] Tom A. B. Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3, 2002.

[138] Tom A. B. Snijders, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.

[139] C.J. Stam. Functional connectivity patterns of human magnetoencephalographic recordings: a 'small-world' network? *Neuroscience Letters*, 355(1):25 – 28, 2004.

[140] Paul E. Stillman, James D. Wilson, Matthew J. Denny, Bruce A. Desmarais, Shankar Bhamidi, Skyler J. Cranmer, and Zhong-Lin Lu. Statistical modeling of the default mode brain network reveals a segregated highway structure. *Scientific Reports*, 7(1):11694, 2017.

[141] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius. Network Analysis of Intrinsic Functional Brain Connectivity in Alzheimer's Disease. *PLoS Computational Biology*, 4:e1000100, June 2008.

[142] Enzo Tagliazucchi, Pablo Balenzuela, Daniel Fraiman, and Dante R Chialvo. Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis. *Frontiers in Physiology*, 3, 2012.

[143] Enzo Tagliazucchi, Frederic Von Wegner, Astrid Morzelewski, Verena Brodbeck, and Helmut Laufs. Dynamic bold functional connectivity in humans and its electrophysiological correlates. *Frontiers in Human Neuroscience*, 6:339, 2012.

[144] Linda S. L. Tan and Nial Friel. Bayesian variational inference for exponential random graph models. *arXiv e-prints*, page arXiv:1811.04249, November 2018.

[145] Jason R. Taylor, Nitin Williams, Rhodri Cusack, Tibor Auer, Meredith A. Shafto, Marie Dixon, Lorraine K. Tyler, Cam-CAN, and Richard N. Henson. The cambridge centre for ageing and neuroscience (cam-can) data repository: Structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 144:262 – 269, 2017. Data Sharing Part II.

[146] S. Thiemichen, N. Friel, A. Caimo, and G. Kauermann. Bayesian exponential random graph models with nodal random effects. *Social Networks*, 46:11–28, July 2016.

[147] William Hedley Thompson, Per Brantefors, and Peter Fransson. From static to temporal network theory: Applications to functional brain connectivity. *Network Neuroscience*, 1(2):69–99, 2017.

[148] Luke Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1728, 12 1994.

[149] Kamen A. Tsvetanov, Richard N.A. Henson, Lorraine K. Tyler, Adeel Razi, Linda Geerligs, Timothy E. Ham, James B. Rowe, and Cam-CAN. Extrinsic and intrinsic brain network connectivity maintains cognition across the lifespan despite accelerated decay of regional brain activation. *Journal of Neuroscience*, 36(11):3115–3126, 2016.

[150] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, 15(1):273 – 289, 2002.

[151] Kâmil Uludağ and Alard Roebroeck. General overview on the merits of multimodal neuroimaging data fusion. *NeuroImage*, 102:3 – 10, 2014. Multimodal Data Fusion.

[152] Martijn P. van den Heuvel, Siemon C. de Lange, Andrew Zalesky, Caio Seguin, B.T. Thomas Yeo, and Ruben Schmidt. Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *NeuroImage*, 152:437 – 449, 2017.

[153] Martijn P. van den Heuvel and Hilleke E. Hulshoff Pol. Exploring the brain network: A review on resting-state fmri functional connectivity. *European Neuropsychopharmacology*, 20(8):519 – 534, 2010.

[154] Martijn P. van den Heuvel and Olaf Sporns. An anatomical substrate for integration among functional networks in human cortex. *Journal of Neuroscience*, 33(36):14489–14500, 2013.

[155] Martijn P. van den Heuvel, Cornelis J. Stam, René S. Kahn, and Hilleke E. Hulshoff Pol. Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, 29(23):7619–7624, 2009.

[156] Bernadette C. M. van Wijk, Cornelis J. Stam, and Andreas Daffertshofer. Comparing brain networks of different size and connectivity density using graph theory. *PLOS ONE*, 5(10):1–13, 10 2010.

[157] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880, Sep. 1997.

[158] Diego Vidaurre, Andrew J. Quinn, Adam P. Baker, David Dupret, Alvaro Tejero-Cantero, and Mark W. Woolrich. Spectrally resolved fast transient brain states in electrophysiological data. *NeuroImage*, 126:81 – 95, 2016.

[159] Huifang E. Wang, Christian G. Bénar, Pascale P. Quilichini, Karl J. Friston, Viktor K. Jirsa, and Christophe Bernard. A systematic framework for functional connectivity measures. *Frontiers in Neuroscience*, 8:405, 2014.

[160] Sumio Watanabe. A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14(Mar):867–897, 2013.

[161] Darren J. Wilkinson and Stephen K. H. Yeung. Conditional simulation from highly structured gaussian systems, with application to blocking-mcmc for the bayesian analysis of very large linear models. *Statistics and Computing*, 12(3):287–300, Jul 2002.

[162] Mark Woolrich, Laurence Hunt, Adrian Groves, and Gareth Barnes. Meg beamforming using bayesian pca for adaptive data covariance matrix regularization. *NeuroImage*, 57(4):1466 – 1479, 2011.

[163] Mark W. Woolrich, Brian D. Ripley, Michael Brady, and Stephen M. Smith. Temporal autocorrelation in univariate linear modeling of {FMRI} data. *NeuroImage*, 14(6):1370 – 1386, 2001.

[164] Mingrui Xia, Jinhui Wang, and Yong He. Brainnet viewer: A network visualization tool for human brain connectomics. *PLOS ONE*, 8(7):1–15, 07 2013.

[165] Yunman Xia, Qunlin Chen, Liang Shi, MengZe Li, Weikang Gong, Hong Chen, and Jiang Qiu. Tracking the dynamic functional connectivity structure of the human brain across the adult lifespan. *Human Brain Mapping*, 40(3):717–728, 2019.

[166] Maziar Yaesoubi, Elena A. Allen, Robyn L. Miller, and Vince D. Calhoun. Dynamic coherence analysis of resting fmri data to jointly capture state-based phase, frequency, and time-domain information. *NeuroImage*, 120:133 – 142, 2015.

[167] Yaming Yu and Xiao-Li Meng. To center or not to center: That is not the question—an ancillarity–sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2011.

[168] Andrew Zalesky, Alex Fornito, and Edward T. Bullmore. Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4):1197 – 1207, 2010.

[169] Andrew Zalesky, Alex Fornito, Luca Cocchi, Leonardo L. Gollo, and Michael Breakspear. Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, 111(28):10341–10346, 2014.